

OM protein - protein search, using sw model

Run on: August 24, 2004, 14:54:57 ; Search time 61.1194 Seconds  
(without alignments)  
69.343 Million cell updates/sec

Title: US-09-641-801-5  
Perfect score: 81  
Sequence: 1 DLEMPVLPVEPFPPFV 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

		%						
Result		Query						
No.	Score	Match	Length	DB	ID		Description	
1	81	100.0	15	4	AAB72504	Aab72504	Colostrin	
2	81	100.0	15	4	AAB59322	Aab59322	Ewe colos	
3	81	100.0	15	4	AAB72250	Aab72250	Colostrin	
4	81	100.0	15	4	AAB72536	Aab72536	Colostrin	
5	81	100.0	15	5	AAO14581	Aao14581	Neural ce	
6	81	100.0	15	5	AAM51040	Aam51040	Colostrin	
7	81	100.0	15	5	AAE20232	Aae20232	Colostrin	
8	81	100.0	16	4	AAB59352	Aab59352	Ewe colos	
9	57	70.4	10	4	AAE07187	Aae07187	Colostrin	

10	51	63.0	10	4	AAE07197	Aae07197	Modified
11	48	59.3	232	7	ADB64957	Adb64957	Human pro
12	45	55.6	55	4	AAU60386	Aau60386	Propionib
13	45	55.6	55	6	ABM56905	Abm56905	Propionib
14	45	55.6	89	3	AAG11334	Aag11334	Arabidops
15	45	55.6	91	3	AAG11333	Aag11333	Arabidops
16	45	55.6	103	3	AAG11332	Aag11332	Arabidops
17	45	55.6	446	6	ABU49258	Abu49258	Protein e
18	45	55.6	619	5	ABP69657	Abp69657	Human pol
19	45	55.6	902	5	ABP65500	Abp65500	Bifidobac
20	44	54.3	247	4	ABG13624	Abg13624	Novel hum
21	44	54.3	259	6	ADA33586	Ada33586	Acinetoba
22	42	51.9	54	4	ABG06766	Abg06766	Novel hum
23	42	51.9	141	6	ABP79310	Abp79310	N. gonorr
24	42	51.9	232	4	ABB60555	Abb60555	Drosophil
25	42	51.9	469	4	ABB61392	Abb61392	Drosophil
26	41.5	51.2	78	3	AAG01794	Aag01794	Human sec
27	41.5	51.2	86	3	AAY86518	Aay86518	Human gen
28	41.5	51.2	86	6	ABO53682	Abo53682	Novel hum
29	41.5	51.2	119	4	AAM91178	Aam91178	Human imm
30	41	50.6	63	4	ABG14725	Abg14725	Novel hum
31	41	50.6	83	4	ABG10974	Abg10974	Novel hum
32	41	50.6	101	4	ABB11472	Abb11472	Human lam
33	41	50.6	133	3	AAG48755	Aag48755	Arabidops
34	41	50.6	153	4	ABB71049	Abb71049	Drosophil
35	41	50.6	159	3	AAG48761	Aag48761	Arabidops
36	41	50.6	159	3	AAG10210	Aag10210	Arabidops
37	41	50.6	174	3	AAG48760	Aag48760	Arabidops
38	41	50.6	174	3	AAG10209	Aag10209	Arabidops
39	41	50.6	193	3	AAG48759	Aag48759	Arabidops
40	41	50.6	194	3	AAG10208	Aag10208	Arabidops
41	41	50.6	220	3	AAG48754	Aag48754	Arabidops
42	41	50.6	235	3	AAG48753	Aag48753	Arabidops
43	41	50.6	318	2	AAY74229	Aay74229	Human pro
44	41	50.6	1421	6	ABO52987	Abo52987	Human put
45	41	50.6	1556	6	ABO53101	Abo53101	Human put

# ALIGNMENTS

## RESULT 1

AAB72504

ID AAB72504 standard; peptide; 15 AA.

XX

AC AAB72504;

XX

DT 09-MAY-2001 (first entry)

XX

DE Colostrinin peptide #5.

XX

KW Dermatological; oxidative stress regulator; colostrinin.

XX

OS Unidentified.

XX

PN WO200112650-A2.

XX

PD 22-FEB-2001.  
 XX  
 PF 17-AUG-2000; 2000WO-US022665.  
 XX  
 PR 17-AUG-1999; 99US-0149310P.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Stanton GJ, Hughes TK, Boldogh I;  
 XX  
 DR WPI; 2001-218342/22.  
 XX  
 PT Modulating oxidative stress level in a cell, involves contacting the cell  
 PT with an oxidative stress regulator selected from colostrinin, its  
 PT constituent peptide, analog or their combinations.  
 XX  
 PS Claim 6; Page 25; 48pp; English.  
 XX  
 CC The present invention relates to a method for modulating the oxidative  
 CC stress level in a cell or a patient, comprising contacting the cell with,  
 CC or administering to the patient, an oxidative stress regulator selected  
 CC from colostrinin, or its constituent peptide (e.g. the present peptide),  
 CC to change the level of an oxidising species in the cell. The method can  
 CC be used to treat oxidative damage to skin, by decreasing or preventing an  
 CC increase in the level of damage to a biomolecule of the patient  
 XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 81; DB 4; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPFPPFV 15  
 |||||  
 Db 1 DLEMPVLPVEPFPPFV 15

# RESULT 2

AAB59322

ID AAB59322 standard; peptide; 15 AA.

XX

AC AAB59322;

XX

DT 21-MAR-2001 (first entry)

XX

DE Ewe colostrinin peptide fragment B-7.

XX

KW Sheep; colostrinin; proline rich polypeptide; colostrum; immune disorder;  
 KW central nervous system disorder; dietary supplement; beta-amyloid plaque.

XX

OS Ovis sp.

XX

PN WO200075173-A2.

XX

PD 14-DEC-2000.

XX

PF 02-JUN-2000; 2000WO-GB002128.

XX  
 PR 02-JUN-1999; 99GB-00012852.  
 XX  
 PA (REGE-) REGEN THERAPEUTICS PLC.  
 XX  
 PI Georgiades JA;  
 XX  
 DR WPI; 2001-071058/08.  
 XX  
 PT Peptides having an N-terminal amino acid sequence isolated from  
 PT colostrinin for treating e.g. disorders of the central nervous system and  
 PT immune system, viral and bacterial infections, and diseases characterized  
 PT by amyloid plaques.  
 XX  
 PS Claim 7; Page 27; 63pp; English.  
 XX  
 CC The present invention provides the sequences of a number of peptides  
 CC found in ewe's colostrinin. Colostrinin is the proline-rich polypeptide  
 CC fragment of colostrum. These peptides can be used in the treatment of  
 CC central nervous system disorders such as senile dementia, Parkinson's  
 CC disease, Alzheimer's disease, psychosis and neurosis, immune system  
 CC disorders such as bacterial and viral infections, to improve the  
 CC development of a child's immune system, as a dietary supplement, and to  
 CC promote the dissolution of beta-amyloid plaques  
 XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 81; DB 4; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEMPVLPVEPFPPFV 15  
 |||||  
 Db 1 DLEMPVLPVEPFPPFV 15

RESULT 3  
 AAB72250  
 ID AAB72250 standard; peptide; 15 AA.  
 XX  
 AC AAB72250;  
 XX  
 DT 14-MAY-2001 (first entry)  
 XX  
 DE Colostrinin derived cytokine inducing peptide SEQ ID 5.  
 XX  
 KW Colostrinin; immune response; cytokine; blood cell proliferation;  
 KW central nervous system disorder; neurological disorder; mental disorder;  
 KW dementia; neurodegenerative disease; Alzheimer's disease; psychosis;  
 KW neurosis; infection.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200111937-A2.  
 XX  
 PD 22-FEB-2001.  
 XX

PF 17-AUG-2000; 2000WO-US022818.

XX

PR 17-AUG-1999; 99US-0149311P.

XX

PA (TEXA ) UNIV TEXAS SYSTEM.

PA (REGE-) REGEN THERAPEUTICS PLC.

XX

PI Stanton GJ, Hughes TK, Boldogh I, Georgiades J;

XX

DR WPI; 2001-202804/20.

XX

PT Inducing a cytokine and modulating an immune response, useful for  
PT treating central nervous system diseases and bacterial and viral  
PT infections, comprises administering colostrinin as an immunological  
PT regulator.

XX

PS Claim 1; Page 34; 50pp; English.

XX

CC Sequences AAB72246 - AAB72275 represent peptides derived from clostrinin,  
CC a proline rich polypeptide aggregate contained in colostrum. The peptides  
CC have immune response modulatory activity, and are capable of inducing  
CC cytokines. Colostrinin and its derived peptides are useful for inducing  
CC cytokine production, for modulating an immunological response and for  
CC inducing blood cell proliferation. The peptides are useful in the  
CC treatment of disorders of the central nervous system, neurological  
CC disorders, mental disorders, dementia, neurodegenerative diseases,  
CC Alzheimer's disease, motor neurone disease, psychosis, neurosis, chronic  
CC disorders of the immune system, bacterial and viral infections and  
CC acquired immunological deficiencies

XX

SQ Sequence 15 AA;

Query Match 100.0%; Score 81; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.1e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLEMPVLPVEPFPPV 15

|||||||

Db 1 DLEMPVLPVEPFPPV 15

#### RESULT 4

AAB72536

ID AAB72536 standard; peptide; 15 AA.

XX

AC AAB72536;

XX

DT 09-MAY-2001 (first entry)

XX

DE Colostrinin peptide #5.

XX

KW Neuroprotective; neural cell differentiation regulator; colostrinin;  
KW colostrum.

XX

OS Unidentified.

XX

PN WO200112651-A2.

XX  
 PD 22-FEB-2001.  
 XX  
 PF 17-AUG-2000; 2000WO-US022774.  
 XX  
 PR 17-AUG-1999; 99US-0149633P.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Boldogh I;  
 XX  
 DR WPI; 2001-226545/23.  
 XX  
 PT Use of colostrinin, its constituent peptide or analog as a neural cell  
 PT regulator, for promoting neural cell differentiation and treating damaged  
 PT neural cells in a patient.  
 XX  
 PS Claim 6; Page 21; 35pp; English.  
 XX  
 CC The present invention relates to a method for promoting neural cell  
 CC differentiation and treating damaged neural cells, using colostrinin and  
 CC colostrinin constituent peptides (e.g. the present peptide) as a neural  
 CC cell regulator. Colostrinin is a polypeptide complex found in colostrum  
 XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 81; DB 4; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPFPPFV 15  
 |||||  
 Db 1 DLEMPVLPVEPFPPFV 15

RESULT 5  
 AAO14581  
 ID AAO14581 standard; peptide; 15 AA.  
 XX  
 AC AAO14581;  
 XX  
 DT 27-MAY-2002 (first entry)  
 XX  
 DE Neural cell regulatory colostrinin peptide 5.  
 XX  
 KW Neural cell differentiation; neural cell regulator; colostrinin peptide;  
 KW neural cell formation; proline-rich polypeptide aggregate; colostrum;  
 KW neural cell treatment.  
 XX  
 OS Unidentified.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 15  
 FT /note= "Optional C-terminal amide"  
 XX  
 PN WO200213851-A1.  
 XX

PD 21-FEB-2002.  
 XX  
 PF 17-AUG-2000; 2000WO-US022777.  
 XX  
 PR 17-AUG-2000; 2000WO-US022777.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Boldogh I, Stanton JG, Hughes TK;  
 XX  
 DR WPI; 2002-269152/31.  
 XX  
 PT Promoting cell differentiation in a patient involves use of blood cell  
 PT regulator selected from colostrinin, its constituent peptide and/or  
 PT analog.  
 XX  
 PS Claim 7; Page 21; 37pp; English.  
 XX  
 CC The invention comprises a method for promoting cell differentiation (e.g.  
 CC neural cell differentiation). The method involves contacting cells with a  
 CC neural cell regulator (i.e. a colostrinin peptide) in order to change the  
 CC cells in morphology to form neural cells. Colostrinin is a proline-rich  
 CC polypeptide aggregate that is present in colostrum. The method of the  
 CC invention is useful for promoting the differentiation of cells and for  
 CC treating damaged neural cells in a patient. The present amino acid  
 CC sequence represents a specifically claimed colostrinin peptide used in  
 CC the method of the invention  
 XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 81; DB 5; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPFPPFV 15  
 |||||  
 Db 1 DLEMPVLPVEPFPPFV 15

# RESULT 6

AAM51040

ID AAM51040 standard; peptide; 15 AA.

XX

AC AAM51040;

XX

DT 30-MAY-2002 (first entry)

XX

DE Colostrinin constituent peptide.

XX

KW Colostrinin; colostrum; immunomodulator; cardiovascular;

KW blood cell regulator; cytokine inducer; beta-casein; human.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Modified-site 15

FT /note= "optional C-terminal amidation"

XX  
 PN WO200213849-A1.  
 XX  
 PD 21-FEB-2002.  
 XX  
 PF 17-AUG-2000; 2000WO-US022775.  
 XX  
 PR 17-AUG-2000; 2000WO-US022775.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 PA (REGE-) REGEN THERAPEUTICS PLC.  
 XX  
 PI Stanton GJ, Hughes TK, Boldogh I, Georgiades J;  
 XX  
 DR WPI; 2002-269150/31.  
 XX  
 PT Modulation of blood cell proliferation in a patient involves use of blood  
 PT cell regulator selected from colostrinin, its constituent peptide and/or  
 PT analog.  
 XX  
 PS Claim 1; Page 34; 54pp; English.  
 XX  
 CC The present sequence is that of a colostrinin constituent peptide that is  
 CC preferred for use as an immunological regulator and as a blood cell  
 CC regulator in claimed methods of the invention. It is classified as having  
 CC a beta-casein homologue precursor. Methods are claimed for: inducing a  
 CC cytokine in a cell by contact with an immunological regulator, where the  
 CC cell is present in a cell culture, a tissue, an organ or an organism, and  
 CC the cell is mammalian, including human; modulating an immune response in  
 CC a cell by contact with the immunological regulator under conditions  
 CC effective to induce a cytokine; modulating an immune response in a  
 CC patient by administering an immunological regulator under conditions  
 CC effective to induce a cytokine, where the immunological regulator is  
 CC administered topically or as part of a dietary supplement, and where the  
 CC immune response is specific or non specific, an interferon response or an  
 CC antibody response; modulating blood cell proliferation by contacting  
 CC blood cells with a blood cell regulator, where the blood cells are  
 CC present in a cell culture or an organism, are mammalian or human, and  
 CC where the blood cells are increased in number or differentiated; and a  
 CC method for modulating blood cell proliferation in a patent. A claimed  
 CC cytokine-inducing composition comprises a pharmaceutical carrier and an  
 CC active agent such as the present peptide. Cytokines induced by this  
 CC peptide in human leucocyte cultures include interferon-gamma, tumour  
 CC necrosis factor-alpha, interleukin-6 and interleukin-10  
 XX  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 81; DB 5; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPFPPFV 15  
 |||||  
 Db 1 DLEMPVLPVEPFPPFV 15

RESULT 7



AAE20232

ID AAE20232 standard; peptide; 15 AA.

XX

AC AAE20232;

XX

DT 18-JUN-2002 (first entry)

XX

DE Colostrinin constituent peptide #5.

XX

KW Blood cell regulator; colostrinin; constituent peptide; oxidative stress;  
KW therapy; oxidative damage; skin; aging; wound healing; cell replacement;  
KW tissue; organ; cosmetic procedure; repair; regeneration; preservation;  
KW transplantation; implantation; dermatological; vulnerary.

XX

OS Unidentified.

XX

FH Key Location/Qualifiers

FT Modified-site 15

FT /note= "Optionally C-terminal amide"

XX

PN WO200213850-A1.

XX

PD 21-FEB-2002.

XX

PF 17-AUG-2000; 2000WO-US022776.

XX

PR 17-AUG-2000; 2000WO-US022776.

XX

PA (TEXA ) UNIV TEXAS SYSTEM.

XX

PI Stanton GJ, Hughes TK, Boldogh I;

XX

DR WPI; 2002-269151/31.

XX

PT Composition useful for the modulation of blood cell proliferation in a  
PT patient comprises a blood cell regulator selected from colostrinin, its  
PT constituent peptide and/or analog.

XX

PS Claim 6; Page 25; 51pp; English.

XX

CC The invention relates to a composition which comprises a blood cell  
CC regulator selected from colostrinin, its constituent peptide and/or  
CC analogue. The invention is used for modulating the oxidative stress level  
CC in a cell e.g. mammalian or human cell present in a cell culture, tissue,  
CC organ, or organism; or for treating oxidative damage to the skin of a  
CC patient e.g. animal or human; to modulate oxidative stress during/ after  
CC a premature birth or normal birth, preventing/delaying aging in a  
CC patient, enhancing wound healing, and the reduction of side effects of  
CC cosmetic procedures. The method changes the level of an oxidising species  
CC in the cell, such as decreases or prevents increase in the level of  
CC damage to a biomolecule of the patient selected from DNA, protein and/or  
CC lipid, compared to the same conditions when the oxidative stress  
CC regulator is not present. The modulation of oxidative stress results in  
CC enhanced repair, regeneration, and replacement of cells, tissues and  
CC organs (e.g. kidney, liver, pancreas, skin, and the other internal and  
CC external organs), as well as enhanced preservation of such organs for  
CC transplantation, implantation, or scientific research. The present

CC sequence is a colostrinin constituent peptide  
XX  
SQ Sequence 15 AA;

Query Match 100.0%; Score 81; DB 5; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.1e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPFPPFV 15  
|||||||  
Db 1 DLEMPVLPVEPFPPFV 15

RESULT 8

AAB59352

ID AAB59352 standard; peptide; 16 AA.

XX

AC AAB59352;

XX

DT 21-MAR-2001 (first entry)

XX

DE Ewe colostrinin peptide fragment derived sequence #12.

XX

KW Sheep; colostrinin; proline rich polypeptide; colostrum; immune disorder;  
KW central nervous system disorder; dietary supplement; beta-amyloid plaque.

XX

OS Ovis sp.

XX

PN WO200075173-A2.

XX

PD 14-DEC-2000.

XX

PF 02-JUN-2000; 2000WO-GB002128.

XX

PR 02-JUN-1999; 99GB-00012852.

XX

PA (REGE-) REGEN THERAPEUTICS PLC.

XX

PI Georgiades JA;

XX

DR WPI; 2001-071058/08.

XX

PT Peptides having an N-terminal amino acid sequence isolated from  
PT colostrinin for treating e.g. disorders of the central nervous system and  
PT immune system, viral and bacterial infections, and diseases characterized  
PT by amyloid plaques.

XX

PS Claim 8; Page 27; 63pp; English.

XX

CC The present invention provides the sequences of a number of peptides  
CC found in ewe's colostrinin. Colostrinin is the proline-rich polypeptide  
CC fragment of colostrum. These peptides can be used in the treatment of  
CC central nervous system disorders such as senile dementia, Parkinson's  
CC disease, Alzheimer's disease, psychosis and neurosis, immune system  
CC disorders such as bacterial and viral infections, to improve the  
CC development of a child's immune system, as a dietary supplement, and to  
CC promote the dissolution of beta-amyloid plaques

XX

SQ Sequence 16 AA;

Query Match 100.0%; Score 81; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 1.2e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPFPPFV 15  
| | | | | | | | | | | | | | |  
Db 2 DLEMPVLPVEPFPPFV 16

RESULT 9

AAE07187

ID AAE07187 standard; peptide; 10 AA.

XX

AC AAE07187;

XX

DT 06-NOV-2001 (first entry)

XX

DE Colostrinin peptide 3.

XX

KW Colostrinin; nootropic; neuroprotective; immunomodulatory; antibacterial;  
KW Parkinson's disease; Alzheimer's disease; mental disorder; food additive;  
KW central nervous system disorder; neurodegenerative disorder; weight loss;  
KW beta-amyloid plaque; psychosis; neurosis; cachexia; motor neuron disease;  
KW acquired immunological deficiency; neurological disorder; dementia;  
KW antiviral.

XX

OS Unidentified.

XX

PN WO200155199-A1.

XX

PD 02-AUG-2001.

XX

PF 26-JAN-2001; 2001WO-GB000329.

XX

PR 26-JAN-2000; 2000GB-00001825.

XX

PA (REGE-) REGEN THERAPEUTICS PLC.

XX

PI Georgiades JA;

XX

DR WPI; 2001-488775/53.

XX

PT Peptide useful as an interalia in the treatment of e.g. disorders of the  
PT immune system and the central nervous system comprises ten amino-terminal  
PT amino acid sequence derived from peptides present in colostrinin.

XX

PS Claim 1; Page 15; 40pp; English.

XX

CC The invention relates to colostrinin peptide fragments which are useful,  
CC inter alia, in the treatment of chronic disorders of the immune system  
CC and the central nervous system. Colostrinin peptides are used as a  
CC medicament in the treatment of neurological disorders e.g., dementia,  
CC neurodegenerative disorders e.g., Alzheimer's disease, motor neuron  
CC disease e.g., Parkinson's disease, mental disorders e.g. psychosis and

CC neurosis, in acquired immunological deficiencies, chronic bacterial and  
 CC viral infections and diseases characterised by the presence of beta-  
 CC amyloid plaques and as a dietary supplement for babies, small children,  
 CC adults and senile persons, who have been subjected to chemotherapy or  
 CC have suffered from cachexia or weight loss due to the chronic disease.  
 CC Colostrinin peptides are also used as food additives and as an auxillary  
 CC withdrawal treatment for drug addicts, after a period of detoxification  
 CC and in persons dependent on stimulants. Colostrinin peptides are used to  
 CC prepare antibodies and to treat emotional disturbances, e.g. emotional  
 CC disturbances of psychiatric patients in a state of depression. These  
 CC colostrinin peptides improves the development of immune system in a new  
 CC born child and to correct the immunological deficiencies in a child. The  
 CC present sequence is colostrinin peptide 3 related to the invention

XX

SQ Sequence 10 AA;

Query Match 70.4%; Score 57; DB 4; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.031;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 PVLPVEPFPPF 14  
 |||||  
 Db 1 PVLPVEPFPPF 10

RESULT 10

AAE07197

ID AAE07197 standard; peptide; 10 AA.

XX

AC AAE07197;

XX

DT 06-NOV-2001 (first entry)

XX

DE Modified colostrinin cyclic peptide #3.

XX

KW Colostrinin; nootropic; neuroprotective; immunomodulatory; antibacterial;  
 KW Parkinson's disease; Alzheimer's disease; mental disorder; food additive;  
 KW central nervous system disorder; neurodegenerative disorder; weight loss;  
 KW beta-amyloid plaque; psychosis; neurosis; cachexia; motor neuron disease;  
 KW acquired immunological deficiency; neurological disorder; dementia;  
 KW antiviral; cyclic.

XX

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Modified-site 1

FT /note= "N-terminal acetyl; this residue forms a cyclic  
 FT linkage with Pro found at the C-terminal end"

XX

PN WO200155199-A1.

XX

PD 02-AUG-2001.

XX

PF 26-JAN-2001; 2001WO-GB000329.

XX

PR 26-JAN-2000; 2000GB-00001825.

XX

PA (REGE-) REGEN THERAPEUTICS PLC.

XX

PI Georgiades JA;

XX

DR WPI; 2001-488775/53.

XX

PT Peptide useful as an interalia in the treatment of e.g. disorders of the  
PT immune system and the central nervous system comprises ten amino-terminal  
PT amino acid sequence derived from peptides present in colostrinin.

XX

PS Example 2; Page 8; 40pp; English.

XX

CC The invention relates to colostrinin peptide fragments which are useful,  
CC inter alia, in the treatment of chronic disorders of the immune system  
CC and the central nervous system. Colostrinin peptides are used as a  
CC medicament in the treatment of neurological disorders e.g., dementia,  
CC neurodegenerative disorders e.g., Alzheimer's disease, motor neuron  
CC disease e.g., Parkinson's disease, mental disorders e.g. psychosis and  
CC neurosis, in acquired immunological deficiencies, chronic bacterial and  
CC viral infections and diseases characterised by the presence of beta-  
CC amyloid plaques and as a dietary supplement for babies, small children,  
CC adults and senile persons, who have been subjected to chemotherapy or  
CC have suffered from cachexia or weight loss due to the chronic disease.  
CC Colostrinin peptides are also used as food additives and as an auxillary  
CC withdrawal treatment for drug addicts, after a period of detoxification  
CC and in persons dependent on stimulants. Colostrinin peptides are used to  
CC prepare antibodies and to treat emotional disturbances, e.g. emotional  
CC disturbances of psychiatric patients in a state of depression. These  
CC colostrinin peptides improves the development of immune system in a new  
CC born child and to correct the immunological deficiencies in a child. The  
CC present sequence is modified colostrinin cyclic peptide #3 related to the  
CC invention

XX

SQ Sequence 10 AA;

Query Match 63.0%; Score 51; DB 4; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.26;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 PVLPVEPFP 13

|||||||

Db 2 PVLPVEPFP 10

RESULT 11

ADB64957

ID ADB64957 standard; protein; 232 AA.

XX

AC ADB64957;

XX

DT 04-DEC-2003 (first entry)

XX

DE Human protein encoded by clone PROST20054660.

XX

KW Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;

KW cell regeneration; membrane protein; signal transduction-related protein;

KW transcription-related protein; osteoporosis; neurological disease;

KW cancer; tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1308459-A2.  
 XX  
 PD 07-MAY-2003.  
 XX  
 PF 28-MAR-2002; 2002EP-00007401.  
 XX  
 PR 05-NOV-2001; 2001JP-00379298.  
 PR 25-JAN-2002; 2002US-00350978.  
 XX  
 PA (HELI-) HELIX RES INST.  
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX  
 PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
 XX  
 DR WPI; 2003-450961/43.  
 DR N-PSDB; ADB62987.  
 XX  
 PT New polynucleotides and polypeptides, useful for developing a diagnostic  
 PT marker or medicines for regulation of their expression and activity, or  
 PT as targets of gene therapy.  
 XX  
 PS Claim 1; Page; 222pp; English.  
 XX  
 CC The invention discloses a polynucleotide comprising a sequence selected  
 CC from 1970 fully defined nucleotide sequences which encode novel  
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
 CC or its partial peptide, an antibody binding to the polypeptide or peptide  
 CC of the polynucleotide, immunologically assaying the polypeptide or  
 CC peptide of the polynucleotide by contacting the polypeptide or peptide  
 CC with the antibody of the encoded protein, and observing the binding  
 CC between the two, a transformant carrying the polynucleotide in an  
 CC expressible manner and an antisense polynucleotide. The oligonucleotide  
 CC is useful as a primer for synthesising the polynucleotide, or as a probe  
 CC for detecting the polynucleotide. The polynucleotides and encoded  
 CC proteins are useful as pharmaceutical agents and many disease-related  
 CC genes may be included in them, for developing a diagnostic marker or  
 CC medicines for regulation of their expression and activity, or as targets  
 CC of gene therapy. The genes are involved in tissue and/or cell  
 CC regeneration. Membrane proteins, signal transduction-related proteins,  
 CC transcription-related proteins, disease-related proteins and genes  
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
 CC neurological diseases, cancer, tumours. The cDNA may be used to regulate  
 CC the activity or expression of the encoded protein to treat diseases. The  
 CC sequence presented is a protein of the invention. Note: Some of the  
 CC sequence data for this patent is not represented in the printed  
 CC specification, but is based on sequence information supplied by the  
 CC European Patent Office.  
 XX  
 SQ Sequence 232 AA;

Query Match

59.3%; Score 48; DB 7; Length 232;

Best Local Similarity 66.7%; Pred. No. 21;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 EMPVLPVEFPF 14  
: ||||| |:||  
Db 209 KFPVLPVHPWPF 220

RESULT 12

AAU60386

ID AAU60386 standard; protein; 55 AA.

XX

AC AAU60386;

XX

DT 27-FEB-2002 (first entry)

XX

DE Propionibacterium acnes immunogenic protein #21282.

XX

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;

KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;

KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;

KW dermatological; osteopathic; neuroprotectant.

XX

OS Propionibacterium acnes.

XX

PN WO200181581-A2.

XX

PD 01-NOV-2001.

XX

PF 20-APR-2001; 2001WO-US012865.

XX

PR 21-APR-2000; 2000US-0199047P.

PR 02-JUN-2000; 2000US-0208841P.

PR 07-JUL-2000; 2000US-0216747P.

XX

PA (CORI-) CORIXA CORP.

XX

PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX

DR WPI; 2001-616774/71.

DR N-PSDB; AAS59609.

XX

PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris.

XX

PS Example 1; SEQ ID NO 21581; 1069pp; English.

XX

CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the

CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 55 AA;

Query Match 55.6%; Score 45; DB 4; Length 55;  
Best Local Similarity 63.6%; Pred. No. 13;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 MPVLPVEPFPPF 14  
:|||| |||:  
Db 23 LPVLPQSPFPY 33

RESULT 13

ABM56905

ID ABM56905 standard; protein; 55 AA.

XX

AC ABM56905;

XX

DT 20-OCT-2003 (first entry)

XX

DE Propionibacterium acnes predicted ORF-encoded polypeptide #21581.

XX

KW Acne vulgaris; antiseborrhoeic; dermatological; antibacterial;

KW immunostimulant; immune response; vaccine.

XX

OS Propionibacterium acnes.

XX

PN WO2003033515-A1.

XX

PD 24-APR-2003.

XX

PF 11-OCT-2002; 2002WO-US032727.

XX

PR 15-OCT-2001; 2001US-00978825.

XX

PA (CORI-) CORIXA CORP.

XX

PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;

PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;

PI Barth B, Vallieve-Douglass J;

XX

DR WPI; 2003-381789/36.

DR N-PSDB; ACF64538.

XX

PT New Propionibacterium acnes polypeptides and polynucleotides encoding the



PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
PT or for stimulating an immune response specific for a P. acnes protein.

XX

PS Example 1; SEQ ID NO 21581; 1481pp; English.

XX

CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
CC encoding a Propionibacterium acnes protein. The invention also relates to  
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
CC immunogenic fragments of P. acnes polypeptides. The invention  
CC additionally encompasses expression vectors and host cells comprising a  
CC polynucleotide of the invention; antibodies against polypeptides of the  
CC invention; fusion proteins comprising a polypeptide of the invention; a  
CC method for stimulating an immune response specific for a P. acnes  
CC polypeptide and an isolated T cell population comprising T cells prepared  
CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
CC antigen-presenting cells that express the polypeptide); a method and kit  
CC for detecting or determining the presence or absence of P. acnes in a  
CC patient; and a method for inhibiting the development of P. acnes in a  
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
CC proteins, T cell populations or antigen-presenting cells that express the  
CC polypeptides are useful for diagnosing, preventing or treating acne  
CC vulgaris, or for stimulating an immune response specific for a P. acnes  
CC protein. The polynucleotides can also be used as probes or primers for  
CC nucleic acid hybridisation. The vaccine composition is useful for the  
CC stimulation of an immune response against P. acnes, or for treating acne,  
CC and the kit is useful for performing a diagnostic assay. The present  
CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
CC reading frame) contained within the P. acnes polynucleotides of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 55 AA;

Query Match 55.6%; Score 45; DB 6; Length 55;  
Best Local Similarity 63.6%; Pred. No. 13;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 MPVLPVEPFPPF 14  
:|||| |||:  
Db 23 LPVLPQSPFPY 33

RESULT 14

AAG11334

ID AAG11334 standard; protein; 89 AA.

XX

AC AAG11334;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 10012.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.

PR	18-JUN-1999;	99US-0139456P.
PR	18-JUN-1999;	99US-0139457P.
PR	18-JUN-1999;	99US-0139458P.
PR	18-JUN-1999;	99US-0139459P.
PR	18-JUN-1999;	99US-0139460P.
PR	18-JUN-1999;	99US-0139461P.
PR	18-JUN-1999;	99US-0139462P.
PR	18-JUN-1999;	99US-0139463P.
PR	18-JUN-1999;	99US-0139750P.
PR	18-JUN-1999;	99US-0139763P.
PR	21-JUN-1999;	99US-0139817P.
PR	22-JUN-1999;	99US-0139899P.
PR	23-JUN-1999;	99US-0140353P.
PR	23-JUN-1999;	99US-0140354P.
PR	24-JUN-1999;	99US-0140695P.
PR	28-JUN-1999;	99US-0140823P.
PR	29-JUN-1999;	99US-0140991P.
PR	30-JUN-1999;	99US-0141287P.
PR	01-JUL-1999;	99US-0141842P.
PR	01-JUL-1999;	99US-0142154P.
PR	02-JUL-1999;	99US-0142055P.
PR	06-JUL-1999;	99US-0142390P.
PR	08-JUL-1999;	99US-0142803P.
PR	09-JUL-1999;	99US-0142920P.
PR	12-JUL-1999;	99US-0142977P.
PR	13-JUL-1999;	99US-0143542P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	16-JUL-1999;	99US-0144086P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.

PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.

PR 21-OCT-1999; 99US-0160767P.  
 PR 21-OCT-1999; 99US-0160768P.  
 PR 21-OCT-1999; 99US-0160770P.  
 PR 21-OCT-1999; 99US-0160814P.  
 PR 21-OCT-1999; 99US-0160815P.  
 PR 22-OCT-1999; 99US-0160980P.  
 PR 22-OCT-1999; 99US-0160981P.  
 PR 22-OCT-1999; 99US-0160989P.  
 PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161406P.  
 PR 26-OCT-1999; 99US-0161359P.  
 PR 26-OCT-1999; 99US-0161360P.  
 PR 26-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161920P.  
 PR 28-OCT-1999; 99US-0161992P.  
 PR 28-OCT-1999; 99US-0161993P.  
 PR 29-OCT-1999; 99US-0162142P.

Query Match 55.6%; Score 45; DB 3; Length 89;  
 Best Local Similarity 77.8%; Pred. No. 22;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 5 PVLPEPEFP 13  
 ||:| ||||  
 Db 47 PVIPTPEFP 55

# RESULT 15

AAG11333

ID AAG11333 standard; protein; 91 AA.

XX

AC AAG11333;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 10011.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-00301439.

XX

PR 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR	06-APR-1999;	99US-0128234P.
PR	08-APR-1999;	99US-0128714P.
PR	16-APR-1999;	99US-0129845P.
PR	19-APR-1999;	99US-0130077P.
PR	21-APR-1999;	99US-0130449P.
PR	23-APR-1999;	99US-0130510P.
PR	23-APR-1999;	99US-0130891P.
PR	28-APR-1999;	99US-0131449P.
PR	30-APR-1999;	99US-0132048P.
PR	30-APR-1999;	99US-0132407P.
PR	04-MAY-1999;	99US-0132484P.
PR	05-MAY-1999;	99US-0132485P.
PR	06-MAY-1999;	99US-0132486P.
PR	06-MAY-1999;	99US-0132487P.
PR	07-MAY-1999;	99US-0132863P.
PR	11-MAY-1999;	99US-0134256P.
PR	14-MAY-1999;	99US-0134218P.
PR	14-MAY-1999;	99US-0134219P.
PR	14-MAY-1999;	99US-0134221P.
PR	14-MAY-1999;	99US-0134370P.
PR	18-MAY-1999;	99US-0134768P.
PR	19-MAY-1999;	99US-0134941P.
PR	20-MAY-1999;	99US-0135124P.
PR	21-MAY-1999;	99US-0135353P.
PR	24-MAY-1999;	99US-0135629P.
PR	25-MAY-1999;	99US-0136021P.
PR	27-MAY-1999;	99US-0136392P.
PR	28-MAY-1999;	99US-0136782P.
PR	01-JUN-1999;	99US-0137222P.
PR	03-JUN-1999;	99US-0137528P.
PR	04-JUN-1999;	99US-0137502P.
PR	07-JUN-1999;	99US-0137724P.
PR	08-JUN-1999;	99US-0138094P.
PR	10-JUN-1999;	99US-0138540P.
PR	10-JUN-1999;	99US-0138847P.
PR	14-JUN-1999;	99US-0139119P.
PR	16-JUN-1999;	99US-0139452P.
PR	16-JUN-1999;	99US-0139453P.
PR	17-JUN-1999;	99US-0139492P.
PR	18-JUN-1999;	99US-0139454P.
PR	18-JUN-1999;	99US-0139455P.
PR	18-JUN-1999;	99US-0139456P.
PR	18-JUN-1999;	99US-0139457P.
PR	18-JUN-1999;	99US-0139458P.
PR	18-JUN-1999;	99US-0139459P.
PR	18-JUN-1999;	99US-0139460P.
PR	18-JUN-1999;	99US-0139461P.
PR	18-JUN-1999;	99US-0139462P.
PR	18-JUN-1999;	99US-0139463P.
PR	18-JUN-1999;	99US-0139750P.
PR	18-JUN-1999;	99US-0139763P.
PR	21-JUN-1999;	99US-0139817P.
PR	22-JUN-1999;	99US-0139899P.
PR	23-JUN-1999;	99US-0140353P.
PR	23-JUN-1999;	99US-0140354P.
PR	24-JUN-1999;	99US-0140695P.
PR	28-JUN-1999;	99US-0140823P.

PR	29-JUN-1999;	99US-0140991P.
PR	30-JUN-1999;	99US-0141287P.
PR	01-JUL-1999;	99US-0141842P.
PR	01-JUL-1999;	99US-0142154P.
PR	02-JUL-1999;	99US-0142055P.
PR	06-JUL-1999;	99US-0142390P.
PR	08-JUL-1999;	99US-0142803P.
PR	09-JUL-1999;	99US-0142920P.
PR	12-JUL-1999;	99US-0142977P.
PR	13-JUL-1999;	99US-0143542P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	16-JUL-1999;	99US-0144086P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.

PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.



PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 55.6%; Score 45; DB 3; Length 91;  
Best Local Similarity 77.8%; Pred. No. 22;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 5 PVLPVEFPF 13  
||:| ||||  
Db 49 PVIPTFPF 57

Search completed: August 24, 2004, 15:42:28  
Job time : 65.1194 secs

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:33:13 ; Search time 16.4552 Seconds  
 (without alignments)  
 47.060 Million cell updates/sec

Title: US-09-641-801-5  
 Perfect score: 81  
 Sequence: 1 DLEMPVLPVEFPFV 15

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
 5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	%		DB	ID	Description
		Match	Length			
1	81	100.0	15	4	US-09-641-803-5	Sequence 5, Appli
2	45	55.6	803	4	US-09-252-991A-30479	Sequence 30479, A
3	44	54.3	259	4	US-09-328-352-4873	Sequence 4873, Ap
4	42	51.9	242	4	US-09-252-991A-28361	Sequence 28361, A
5	41.5	51.2	86	4	US-09-461-325-456	Sequence 456, App
6	41.5	51.2	86	4	US-10-012-542-456	Sequence 456, App
7	40	49.4	143	4	US-09-621-976-5226	Sequence 5226, Ap
8	40	49.4	219	4	US-09-527-345-2	Sequence 2, Appli
9	40	49.4	577	4	US-09-489-039A-9575	Sequence 9575, Ap
10	39	48.1	220	4	US-09-198-452A-211	Sequence 211, App
11	39	48.1	373	4	US-09-149-476-374	Sequence 374, App

12	39	48.1	405	3	US-08-888-429A-22	Sequence 22, Appl
13	39	48.1	405	4	US-09-593-653-22	Sequence 22, Appl
14	39	48.1	444	4	US-09-252-991A-20775	Sequence 20775, A
15	39	48.1	526	1	US-07-921-796-6	Sequence 6, Appli
16	39	48.1	526	1	US-07-921-796-8	Sequence 8, Appli
17	39	48.1	540	4	US-08-945-771-2	Sequence 2, Appli
18	38	46.9	110	4	US-09-543-681A-5498	Sequence 5498, Ap
19	38	46.9	455	4	US-09-489-039A-9964	Sequence 9964, Ap
20	38	46.9	477	3	US-08-704-711A-20	Sequence 20, Appl
21	38	46.9	477	3	US-08-448-489-15	Sequence 15, Appl
22	38	46.9	477	3	US-08-281-313-1	Sequence 9, Appli
23	38	46.9	477	4	US-09-521-220-20	Sequence 20, Appl
24	38	46.9	477	4	US-09-391-104-21	Sequence 21, Appl
25	38	46.9	486	4	US-08-259-451-13	Sequence 13, Appl
26	38	46.9	548	4	US-09-252-991A-28958	Sequence 28958, A
27	38	46.9	655	4	US-09-252-991A-25314	Sequence 25314, A
28	38	46.9	820	4	US-09-252-991A-32001	Sequence 32001, A
29	38	46.9	937	4	US-09-252-991A-19108	Sequence 19108, A
30	38	46.9	959	4	US-09-543-681A-6879	Sequence 6879, Ap
31	37	45.7	367	1	US-07-864-004B-2	Sequence 2, Appli
32	37	45.7	367	1	US-08-251-937A-2	Sequence 2, Appli
33	37	45.7	367	4	US-09-198-452A-1069	Sequence 1069, Ap
34	37	45.7	367	5	PCT-US93-03275-2	Sequence 2, Appli
35	37	45.7	368	1	US-08-212-133A-6	Sequence 6, Appli
36	37	45.7	368	1	US-08-474-503-4	Sequence 4, Appli
37	37	45.7	368	2	US-08-670-707A-4	Sequence 4, Appli
38	37	45.7	368	3	US-09-037-601-4	Sequence 4, Appli
39	37	45.7	368	4	US-09-315-179-4	Sequence 4, Appli
40	37	45.7	368	4	US-09-523-656-4	Sequence 4, Appli
41	37	45.7	368	5	PCT-US94-13200-4	Sequence 4, Appli
42	37	45.7	372	4	US-09-252-991A-22962	Sequence 22962, A
43	37	45.7	429	4	US-09-328-352-4875	Sequence 4875, Ap
44	37	45.7	482	4	US-09-489-039A-7931	Sequence 7931, Ap
45	37	45.7	486	1	US-07-672-483-2	Sequence 2, Appli

# ALIGNMENTS

## RESULT 1

US-09-641-803-5

; Sequence 5, Application US/09641803

; Patent No. 6500798

; GENERAL INFORMATION:

; APPLICANT: STANTON, G. John

; APPLICANT: HUGHES, Thomas K.

; APPLICANT: BOLDOGH, Istvan

; TITLE OF INVENTION: USE OF COLOSTRININ, CONSTITUENT PEPTIDES THEREOF, AND

; TITLE OF INVENTION: ANALOGS THEREOF AS OXIDATIVE STRESS REGULATORS

; FILE REFERENCE: 265.00220101

; CURRENT APPLICATION NUMBER: US/09/641,803

; CURRENT FILING DATE: 2000-08-17

; PRIOR APPLICATION NUMBER: 60/149,310

; PRIOR FILING DATE: 1999-08-17

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 5

; LENGTH: 15  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: peptide  
US-09-641-803-5

Query Match 100.0%; Score 81; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2.8e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPFPPFV 15  
| | | | | | | | | | | | | | |  
Db 1 DLEMPVLPVEPFPPFV 15

RESULT 2

US-09-252-991A-30479  
; Sequence 30479, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30479  
; LENGTH: 803  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30479

Query Match 55.6%; Score 45; DB 4; Length 803;  
Best Local Similarity 53.3%; Pred. No. 50;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPFPPFV 15  
| | | | | : | : |  
Db 318 DAEPPVVPVQVLPYV 332

RESULT 3

US-09-328-352-4873  
; Sequence 4873, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ACINETOBACTER

; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 4873  
; LENGTH: 259  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-4873

Query Match 54.3%; Score 44; DB 4; Length 259;  
Best Local Similarity 58.3%; Pred. No. 21;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 4 MPVLPVEPFPPFV 15  
| | ||:|:|:  
Db 166 MVVRPVPDPYPI 177

RESULT 4

US-09-252-991A-28361  
; Sequence 28361, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 28361  
; LENGTH: 242  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-28361

Query Match 51.9%; Score 42; DB 4; Length 242;  
Best Local Similarity 58.3%; Pred. No. 40;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 3 EMPVLPVEPFPPF 14  
| | | :|||  
Db 60 ESPQRPAQPFPPF 71

RESULT 5

US-09-461-325-456  
; Sequence 456, Application US/09461325A  
; Patent No. 6475753  
; GENERAL INFORMATION:

```
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: PZ029P1
; CURRENT APPLICATION NUMBER: US/09/461,325A
; CURRENT FILING DATE: 1999-12-14
; EARLIER APPLICATION NUMBER: PCT/US99/13418
; EARLIER FILING DATE: 1999-06-15
; EARLIER APPLICATION NUMBER: 60/089,507
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,508
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,509
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/089,510
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: 60/090,112
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: 60/090,113
; EARLIER FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 456
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-325-456
```

```
Query Match          51.2%; Score 41.5; DB 4; Length 86;
Best Local Similarity 47.1%; Pred. No. 16;
Matches      8; Conservative 3; Mismatches 1; Indels 5; Gaps 1;
```

```
Qy      2 LEMPVLP-----VEPFP 13
          ||:|:|      : |||
Db      17 LEVPILPTHLLIHPFP 33
```

# RESULT 6

```
US-10-012-542-456
; Sequence 456, Application US/10012542
; Patent No. 6627741
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: PZ029P1
; CURRENT APPLICATION NUMBER: US/10/012,542
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112
```

; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22  
; NUMBER OF SEQ ID NOS: 532  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 456  
; LENGTH: 86  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-012-542-456

Query Match 51.2%; Score 41.5; DB 4; Length 86;  
Best Local Similarity 47.1%; Pred. No. 16;  
Matches 8; Conservative 3; Mismatches 1; Indels 5; Gaps 1;

Qy 2 LEMPVLP-----VEPFP 13  
||:|:| : |||  
Db 17 LEVPILPTHLLIHPFP 33

RESULT 7

US-09-621-976-5226  
; Sequence 5226, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 5226  
; LENGTH: 143  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -15..-1  
; NAME/KEY: UNSURE  
; LOCATION: 111  
; OTHER INFORMATION: Xaa = Ala,Pro  
US-09-621-976-5226

Query Match 49.4%; Score 40; DB 4; Length 143;  
Best Local Similarity 57.1%; Pred. No. 46;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 2 LEMPVLPVEPFPFV 15  
| :| || |||  
Db 99 LNVPLPPRGFPFV 112

RESULT 8

US-09-527-345-2

```
; Sequence 2, Application US/09527345
; Patent No. 6331413
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/527,345
; CURRENT FILING DATE: 1999-03-17
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-527-345-2
```

```
Query Match          49.4%; Score 40; DB 4; Length 219;
Best Local Similarity 57.1%; Pred. No. 72;
Matches      8; Conservative 1; Mismatches      5; Indels      0; Gaps      0;
```

```
Qy      2 LEMPVLPVEPFPPV 15
        | :| ||  |||
Db      99 LNVPLPLPPRGFPFV 112
```

#### RESULT 9

```
US-09-489-039A-9575
; Sequence 9575, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 9575
; LENGTH: 577
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9575
```

```
Query Match          49.4%; Score 40; DB 4; Length 577;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches      8; Conservative 0; Mismatches      4; Indels      0; Gaps      0;
```

```
Qy      2 LEMPVLPVEPFP 13
        |||  |||  ||
Db      302 LEMDTLPVSPAP 313
```



RESULT 10

US-09-198-452A-211

; Sequence 211, Application US/09198452A  
 ; Patent No. 6559294  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Griffais, R.  
 ; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
 ; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
 ; TITLE OF INVENTION: and treatment of infection  
 ; FILE REFERENCE: 9710-003-999  
 ; CURRENT APPLICATION NUMBER: US/09/198,452A  
 ; CURRENT FILING DATE: 1998-11-24  
 ; NUMBER OF SEQ ID NOS: 6849  
 ; SEQ ID NO 211  
 ; LENGTH: 220  
 ; TYPE: PRT  
 ; ORGANISM: Chlamydia pneumoniae  
 US-09-198-452A-211

Query Match 48.1%; Score 39; DB 4; Length 220;  
 Best Local Similarity 70.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 6 VLPVEPFPPFV 15  
 |||:|| | |  
 Db 82 VLPIEPTPLV 91

RESULT 11

US-09-149-476-374

; Sequence 374, Application US/09149476  
 ; Patent No. 6420526  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: 186 Human Secreted proteins  
 ; FILE REFERENCE: PZ002P1  
 ; CURRENT APPLICATION NUMBER: US/09/149,476  
 ; CURRENT FILING DATE: 1998-09-08  
 ; EARLIER APPLICATION NUMBER: PCT/US98/04493  
 ; EARLIER FILING DATE: 1998-03-06  
 ; EARLIER APPLICATION NUMBER: 60/040,162  
 ; EARLIER FILING DATE: 1997-03-07  
 ; EARLIER APPLICATION NUMBER: 60/040,333  
 ; EARLIER FILING DATE: 1997-03-07  
 ; EARLIER APPLICATION NUMBER: 60/038,621  
 ; EARLIER FILING DATE: 1997-03-07  
 ; EARLIER APPLICATION NUMBER: 60/040,626  
 ; EARLIER FILING DATE: 1997-03-07  
 ; EARLIER APPLICATION NUMBER: 60/040,334  
 ; EARLIER FILING DATE: 1997-03-07  
 ; EARLIER APPLICATION NUMBER: 60/040,336  
 ; EARLIER FILING DATE: 1997-03-07  
 ; EARLIER APPLICATION NUMBER: 60/040,163  
 ; EARLIER FILING DATE: 1997-03-07

; EARLIER APPLICATION NUMBER: 60/047,600  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,615  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,597  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,502  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,633  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,583  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,617  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,618  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,503  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,592  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,581  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,584  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,500  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,587  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,492  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,598  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,613  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,582  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,596  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,612  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,632  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,601  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/043,580  
; EARLIER FILING DATE: 1997-04-11  
; EARLIER APPLICATION NUMBER: 60/043,568  
; EARLIER FILING DATE: 1997-04-11  
; EARLIER APPLICATION NUMBER: 60/043,314  
; EARLIER FILING DATE: 1997-04-11  
; EARLIER APPLICATION NUMBER: 60/043,569  
; EARLIER FILING DATE: 1997-04-11  
; EARLIER APPLICATION NUMBER: 60/043,311  
; EARLIER FILING DATE: 1997-04-11  
; EARLIER APPLICATION NUMBER: 60/043,671  
; EARLIER FILING DATE: 1997-04-11  
; EARLIER APPLICATION NUMBER: 60/043,674

; EARLIER FILING DATE: 1997-04-11  
; EARLIER APPLICATION NUMBER: 60/043,669  
; EARLIER FILING DATE: 1997-04-11  
; EARLIER APPLICATION NUMBER: 60/043,312  
; EARLIER FILING DATE: 1997-04-11  
; EARLIER APPLICATION NUMBER: 60/043,313  
; EARLIER FILING DATE: 1997-04-11  
; EARLIER APPLICATION NUMBER: 60/043,672  
; EARLIER FILING DATE: 1997-04-11  
; EARLIER APPLICATION NUMBER: 60/043,315  
; EARLIER FILING DATE: 1997-04-11  
; EARLIER APPLICATION NUMBER: 60/048,974  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/056,886  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,877  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,889  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,893  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,630  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,878  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,662  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,872  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,882  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,637  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,903  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,888  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,879  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,880  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,894  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,911  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,636  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,874  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,910  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,864  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,631  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,845  
; EARLIER FILING DATE: 1997-08-22

; EARLIER APPLICATION NUMBER: 60/056,892  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/057,761  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/047,595  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,599  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,588  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,585  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,586  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,590  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,594  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,589  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,593  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,614  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/043,578  
; EARLIER FILING DATE: 1997-04-11  
; EARLIER APPLICATION NUMBER: 60/043,576  
; EARLIER FILING DATE: 1997-04-11  
; EARLIER APPLICATION NUMBER: 60/047,501  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/043,670  
; EARLIER FILING DATE: 1997-04-11  
; EARLIER APPLICATION NUMBER: 60/056,632  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,664  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,876  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,881  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,909  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,875  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,862  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,887  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/056,908  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/048,964  
; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/057,650  
; EARLIER FILING DATE: 1997-09-05  
; EARLIER APPLICATION NUMBER: 60/056,884  
; EARLIER FILING DATE: 1997-08-22  
; EARLIER APPLICATION NUMBER: 60/057,669

; EARLIER FILING DATE: 1997-09-05  
; EARLIER APPLICATION NUMBER: 60/049,610  
; EARLIER FILING DATE: 1997-06-13  
; EARLIER APPLICATION NUMBER: 60/061,060  
; EARLIER FILING DATE: 1997-10-02

Query Match 48.1%; Score 39; DB 4; Length 373;  
Best Local Similarity 41.7%; Pred. No. 1.8e+02;  
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPF 12  
::||| |::||  
Db 104 EMEVPQAPIQPF 115

RESULT 12

US-08-888-429A-22

; Sequence 22, Application US/08888429A  
; Patent No. 6136596

; GENERAL INFORMATION:

; APPLICANT: Davis, Roger J.  
; APPLICANT: Whitmarsh, Alan  
; APPLICANT: Tournier, Cathy  
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND ONCOPROTEIN-  
; TITLE OF INVENTION: ACTIVATED HUMAN PROTEIN KINASE KINASES  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/888,429A  
; FILING DATE: 07-JUL-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/530,950  
; FILING DATE: 19-SEP-1995  
; APPLICATION NUMBER: 08/446,083  
; FILING DATE: 19-MAY-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Fasse, Peter J.  
; REGISTRATION NUMBER: 32,983  
; REFERENCE/DOCKET NUMBER: 07917/053001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 299354

; INFORMATION FOR SEQ ID NO: 22:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 405 amino acids

; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-888-429A-22

Query Match 48.1%; Score 39; DB 3; Length 405;  
Best Local Similarity 58.8%; Pred. No. 2e+02;  
Matches 10; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

Qy 1 DLEMPVLPVEPF--PFV 15  
| : |||| | ||  
Db 328 DEDSPVLPVGEFSEPFV 344

RESULT 13

US-09-593-653-22

; Sequence 22, Application US/09593653  
; Patent No. 6610523

; GENERAL INFORMATION:

; APPLICANT: Davis, Roger J.  
; Whitmarsh, Alan  
; Tournier, Cathy

; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND ONCOPROTEIN-  
; ACTIVATED HUMAN PROTEIN KINASE KINASES

; NUMBER OF SEQUENCES: 34

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/593,653  
; FILING DATE: 13-Jun-2000

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/888,429A  
; FILING DATE: 07-JUL-1997  
; APPLICATION NUMBER: 08/530,950  
; FILING DATE: 19-SEP-1995  
; APPLICATION NUMBER: 08/446,083  
; FILING DATE: 19-MAY-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Fasse, Peter J.  
; REGISTRATION NUMBER: 32,983  
; REFERENCE/DOCKET NUMBER: 07917/053001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 299354

; INFORMATION FOR SEQ ID NO: 22:

; SEQUENCE CHARACTERISTICS:

;           LENGTH: 405 amino acids  
;           TYPE: amino acid  
;           TOPOLOGY: linear  
;       MOLECULE TYPE: protein  
;       SEQUENCE DESCRIPTION: SEQ ID NO: 22:  
US-09-593-653-22

Query Match           48.1%;   Score 39;   DB 4;   Length 405;  
Best Local Similarity   58.8%;   Pred. No. 2e+02;  
Matches   10;   Conservative   1;   Mismatches   4;   Indels    2;   Gaps    1;

Qy           1 DLEMPVLPVEPF--PFV 15  
              | : ||||| | |||  
Db           328 DEDSPVLPVGEFSEPFV 344

RESULT 14

US-09-252-991A-20775  
; Sequence 20775, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 20775  
;   LENGTH: 444  
;   TYPE: PRT  
;   ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20775

Query Match           48.1%;   Score 39;   DB 4;   Length 444;  
Best Local Similarity   53.8%;   Pred. No. 2.2e+02;  
Matches    7;   Conservative   1;   Mismatches   5;   Indels    0;   Gaps    0;

Qy           2 LEMPVLPVEPFPPF 14  
              | :|| | |||  
Db           62 LALPVCPCRPSPF 74

RESULT 15

US-07-921-796-6  
; Sequence 6, Application US/07921796  
; Patent No. 5487990  
; GENERAL INFORMATION:  
; APPLICANT: Smith, John A.  
; APPLICANT: Lee, Fang-Jen S.  
; APPLICANT: Lin, Lee-Wen  
; TITLE OF INVENTION: The Glucose-Regulated Promoter of Yeast

```

; TITLE OF INVENTION: Acetyl-CoA Hydrolase
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/921,796
; FILING DATE: 30-JUL-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanzo, Michael A.
; REGISTRATION NUMBER: 36,912
; REFERENCE/DOCKET NUMBER: 0609.1600003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 526 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-921-796-6

```

```

Query Match          48.1%; Score 39; DB 1; Length 526;
Best Local Similarity 41.2%; Pred. No. 2.6e+02;
Matches      7; Conservative      6; Mismatches      2; Indels      2; Gaps      1;

```

```

QY      1 DLEMPVLPV--EPFPFV 15
        |::||| |   :|:|::
Db      191 DIDMPVNPPFRKPYPYL 207

```

```

Search completed: August 24, 2004, 15:55:15
Job time : 17.4552 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:26:28 ; Search time 14.5522 Seconds  
(without alignments)  
99.151 Million cell updates/sec

Title: US-09-641-801-5  
Perfect score: 81  
Sequence: 1 DLEMPVLPVEPFPPFV 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pirl:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query Match	Length			
1	45	55.6	219	2	T24429	hypothetical prote
2	45	55.6	446	2	H82264	probable capK prot
3	43	53.1	421	2	H69490	formylmethanofuran
4	43	53.1	556	2	T24690	hypothetical prote
5	42	51.9	218	2	S54532	probable membrane
6	42	51.9	295	4	JC6530	laminin receptor p
7	42	51.9	367	2	T24298	hypothetical prote
8	42	51.9	448	2	AF1942	hypothetical prote
9	42	51.9	473	2	AF2433	aldehyde dehydroge
10	42	51.9	696	2	T07447	DNA-directed RNA p
11	42	51.9	1504	2	T17426	FK506 polyketide s
12	41	50.6	503	2	B97432	glucose-6-phosphat
13	41	50.6	973	2	T41201	isoleucyl-trna syn

14	41	50.6	1421	2	T00333	hypothetical prote
15	40	49.4	266	2	AG3000	1-acyl-sn-glycerol
16	40	49.4	266	2	B98283	hypothetical prote
17	40	49.4	360	2	S75350	hypothetical prote
18	40	49.4	435	2	T37003	probable flavohemo
19	40	49.4	599	2	AB1906	hypothetical prote
20	40	49.4	727	2	AD1868	hypothetical prote
21	40	49.4	1161	1	S31213	nidogen precursor
22	39.5	48.8	1086	2	JC7736	C3G protein, long
23	39	48.1	207	1	S06869	hypothetical prote
24	39	48.1	217	2	H86514	ABC amino acid tra
25	39	48.1	217	2	G72108	amino acid ABC tra
26	39	48.1	333	2	G75055	hypothetical prote
27	39	48.1	344	2	C95884	probable oxidoredu
28	39	48.1	347	2	T05737	probable hordein C
29	39	48.1	361	2	T46704	hypothetical prote
30	39	48.1	415	2	T00614	hypothetical prote
31	39	48.1	444	2	A56708	MAPK/ERK kinase 5
32	39	48.1	526	1	S28549	acetyl-CoA hydroly
33	39	48.1	739	2	T45429	polyphosphate kina
34	39	48.1	742	2	E70673	probable ppk prote
35	39	48.1	1020	2	T18260	1-phosphatidylinos
36	38.5	47.5	302	2	S21357	beta-casein - tamm
37	38.5	47.5	413	2	AH2743	conserved hypothet
38	38.5	47.5	416	2	G97524	hypothetical prote
39	38.5	47.5	431	2	S37775	filamin, muscle -
40	38	46.9	52	2	E98316	hypothetical prote
41	38	46.9	166	2	D75173	hypothetical prote
42	38	46.9	190	2	A69302	hypothetical prote
43	38	46.9	300	2	S27780	major merozoite su
44	38	46.9	344	2	C84015	transcription regu
45	38	46.9	369	2	G83351	probable ATP-bind

# ALIGNMENTS

## RESULT 1

T24429

hypothetical protein T04A8.11 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T24429

R;Palmer, S.

submitted to the EMBL Data Library, August 1994

A;Reference number: Z19889

A;Accession: T24429

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-219 <WIL>

A;Cross-references: EMBL:Z35663; PIDN:CAA84730.1; GSPDB:GN00021; CESP:T04A8.11

A;Experimental source: clone T04A8

C;Genetics:

A;Gene: CESP:T04A8.11

A;Map position: 3

A;Introns: 12/1; 40/2; 148/3

Query Match 55.6%; Score 45; DB 2; Length 219;  
Best Local Similarity 63.6%; Pred. No. 5.9;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LEMPVLPVEPF 12  
|::||:| |||  
Db 21 LKLPVMPAEPF 31

RESULT 2

H82264

probable capK protein VC0924 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)

C;Species: *Vibrio cholerae*

C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C;Accession: H82264

R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Umayam, L.A.; Gill, S.R.; Nelson, K.E.; Read, T.D.; Tettelin, H.; Richardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.; McDonald, L.; Utterback, T.; Fleishmann, R.D.; Nierman, W.C.; White, O.; Salzberg, S.L.; Smith, H.O.; Colwell, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M. Nature 406, 477-483, 2000

A;Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.

A;Reference number: A82035; MUID:20406833; PMID:10952301

A;Accession: H82264

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-446 <HEI>

A;Cross-references: GB:AE004175; GB:AE003852; NID:g9655366; PIDN:AAF94086.1; GSPDB:GN00126; TIGR:VC0924

A;Experimental source: serogroup O1; strain N16961; biotype El Tor

C;Genetics:

A;Gene: VC0924

A;Map position: 1

Query Match 55.6%; Score 45; DB 2; Length 446;  
Best Local Similarity 53.3%; Pred. No. 13;  
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPFPPFV 15  
|||:| | :| ||::  
Db 76 DLEVPNLELEAFPYL 90

RESULT 3

H69490

formylmethanofuran dehydrogenase (tungsten) chain B homolog (fwdB-2) - *Archaeoglobus fulgidus*

C;Species: *Archaeoglobus fulgidus*

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Sep-1999

C;Accession: H69490

R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny,

K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*.  
 A;Reference number: A69250; MUID:98049343; PMID:9389475  
 A;Accession: H69490  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-421 <KLE>  
 A;Cross-references: GB:AE000970; GB:AE000782; NID:g2689293; PIDN:AAB89326.1; PID:g2648615; TIGR:AF1929  
 C;Superfamily: formylmethanofuran dehydrogenase (molybdenum) beta chain  
 C;Keywords: iron-sulfur protein; metalloprotein; molybdenum; molybdopterin

Query Match 53.1%; Score 43; DB 2; Length 421;  
 Best Local Similarity 54.5%; Pred. No. 26;  
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 EMPVLPVEPF 13  
 |::: ::|||  
 Db 350 EIPVIQIDPF 360

#### RESULT 4

T24690  
 hypothetical protein T08D2.8 - *Caenorhabditis elegans*  
 C;Species: *Caenorhabditis elegans*  
 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C;Accession: T24690  
 R;McMurray, A.  
 submitted to the EMBL Data Library, March 1997  
 A;Reference number: Z19924  
 A;Accession: T24690  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-556 <WIL>  
 A;Cross-references: EMBL:Z92839; PIDN:CAB07415.1; CESP:T08D2.8  
 A;Experimental source: clone T08D2  
 C;Genetics:  
 A;Gene: CESP:T08D2.8  
 A;Introns: 5/1; 54/3; 80/3; 255/3; 307/3; 352/3; 479/1

Query Match 53.1%; Score 43; DB 2; Length 556;  
 Best Local Similarity 72.7%; Pred. No. 36;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 PVLPEPFPPFV 15  
 || ||:| |||  
 Db 443 PVAPVKPKPFV 453

#### RESULT 5

S54532  
probable membrane protein YDR236c - yeast (*Saccharomyces cerevisiae*)  
N;Alternate names: hypothetical protein YD8419.03c  
C;Species: *Saccharomyces cerevisiae*  
C;Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 19-Apr-2002  
C;Accession: S54532  
R;Oliver, K.; Harris, D.  
submitted to the EMBL Data Library, May 1995  
A;Reference number: S54530  
A;Accession: S54532  
A;Molecule type: DNA  
A;Residues: 1-218 <OLI>  
A;Cross-references: EMBL:Z49701; NID:g817819; PID:g817822; GSPDB:GN00004;  
MIPS:YDR236c  
A;Experimental source: strain AB972  
C;Genetics:  
A;Gene: SGD:FMN1; MIPS:YDR236c  
A;Cross-references: SGD:S0002644  
A;Map position: 4R  
C;Keywords: transmembrane protein  
F;4-20/Domain: transmembrane #status predicted <TMM>

Query Match 51.9%; Score 42; DB 2; Length 218;  
Best Local Similarity 53.3%; Pred. No. 18;  
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPFPPFV 15  
|| :| | ||| |  
Db 37 DLPIPAQPGPPFPLV 51

# RESULT 6

JC6530  
laminin receptor processed pseudogene LAMRL5 - human  
C;Species: *Homo sapiens* (man)  
C;Date: 28-Aug-1998 #sequence\_revision 28-Aug-1998 #text\_change 28-Aug-1998  
C;Accession: JC6530  
R;Richardson, M.P.; Braybrook, C.; Tham, M.; Moore, G.E.; Stanier, P.  
Gene 206, 145-150, 1998  
A;Title: Molecular cloning and characterization of a highly conserved human 67-  
kDa laminin receptor pseudogene mapping to Xq21.3.  
A;Reference number: JC6530; MUID:98121324; PMID:9461426  
A;Accession: JC6530  
A;Status: conceptual translation of pseudogene  
A;Molecule type: DNA  
A;Residues: 1-295 <RIC>  
A;Experimental source: brain  
C;Comment: No evidence could be found that this intronless gene sequence is  
expressed.  
C;Genetics:  
A;Gene: LAMRL5  
A;Map position: Xq21.3  
A;Introns: #status absent  
C;Keywords: brain; glycoprotein; laminin binding; pseudogene; receptor

Query Match 51.9%; Score 42; DB 4; Length 295;  
Best Local Similarity 46.2%; Pred. No. 26;

Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPFP 13  
::|:| :|:| ||  
Db 251 EVEVPSVPIEEFP 263

RESULT 7

T24298  
hypothetical protein T01E8.2 - *Caenorhabditis elegans*  
C;Species: *Caenorhabditis elegans*  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T24298  
R;McMurray, A.  
submitted to the EMBL Data Library, March 1995  
A;Reference number: Z19871  
A;Accession: T24298  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-367 <WIL>  
A;Cross-references: EMBL:Z48809; PIDN:CAA88744.1; GSPDB:GN00020; CESP:T01E8.2  
A;Experimental source: clone T01E8  
C;Genetics:  
A;Gene: CESP:T01E8.2  
A;Map position: 2  
A;Introns: 48/2; 200/2; 254/3

Query Match 51.9%; Score 42; DB 2; Length 367;  
Best Local Similarity 60.0%; Pred. No. 33;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 5 PVLPVEPFPF 14  
|: | |||  
Db 295 PIFPFRPFPF 304

RESULT 8

AF1942  
hypothetical protein all1089 [imported] - *Nostoc* sp. (strain PCC 7120)  
C;Species: *Nostoc* sp. PCC 7120  
A;Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C;Accession: AF1942  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.;  
Iriguchi, M.; Ishikawa, A.; Kawashima, K.; Kimura, T.; Kishida, Y.; Kohara, M.;  
Matsumoto, M.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.;  
Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing  
*Cyanobacterium Anabaena* sp. strain PCC 7120.  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AF1942  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-448 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BA073046.1; PID:g17130435; GSPDB:GN00179  
A;Experimental source: strain PCC 7120

C;Genetics:  
A;Gene: all1089

Query Match 51.9%; Score 42; DB 2; Length 448;  
Best Local Similarity 60.0%; Pred. No. 41;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 3 EMPVLPVEPF 12  
|:|:|:|:  
Db 98 ELPLLPVDPY 107

RESULT 9  
AF2433

aldehyde dehydrogenase [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp. PCC 7120  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C;Accession: AF2433  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.;  
Iriguchi, M.; Ishikawa, A.; Kawashima, K.; Kimura, T.; Kishida, Y.; Kohara, M.;  
Matsumoto, M.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.;  
Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing  
Cyanobacterium Anabaena sp. strain PCC 7120.  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AF2433  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-473 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BAB76721.1; PID:g17134160; GSPDB:GN00179  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: all15022  
C;Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 51.9%; Score 42; DB 2; Length 473;  
Best Local Similarity 63.6%; Pred. No. 44;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 5 PVLPVEPFPPFV 15  
|:|:| | | | |  
Db 365 PIMPVMPFPDV 375

RESULT 10  
T07447

DNA-directed RNA polymerase (EC 2.7.7.6) beta'-1 chain - Japanese black pine  
chloroplast (fragment)  
C;Species: chloroplast Pinus thunbergiana (Japanese black pine)  
C;Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 18-Aug-2000  
C;Accession: T07447  
R;Wakasugi, T.; Tsudzuki, J.; Ito, S.; Nakashima, K.; Tsudzuki, T.; Sugiura, M.  
Proc. Natl. Acad. Sci. U.S.A. 91, 9794-9798, 1994  
A;Title: Loss of all ndh genes as determined by sequencing the entire  
chloroplast genome of the black pine Pinus thunbergii.

A;Reference number: Z16030; MUID:95024047; PMID:7937893  
A;Accession: T07447  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-696 <WAK>  
A;Cross-references: EMBL:D17510; NID:g529643; PIDN:BAA23472.1; PID:g2626945  
C;Genetics:  
A;Gene: rpoC1  
A;Genome: chloroplast  
A;Note: intron positions not resolved (incomplete sequence)  
C;Superfamily: chloroplast DNA-directed RNA polymerase beta'-1 chain  
C;Keywords: chloroplast; nucleotidyltransferase; transcription

Query Match 51.9%; Score 42; DB 2; Length 696;  
Best Local Similarity 66.7%; Pred. No. 68;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 MPVLPVEPPFPFV 15  
:|||| || | |  
Db 285 LPVLPPEPRPIV 296

RESULT 11

T17426

FK506 polyketide synthetase fkbP [imported] - Streptomyces sp. (strain MA6548)

C;Species: Streptomyces sp.

A;Variety: strain MA6548

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 03-Nov-2000

C;Accession: T17426

R;Motamedi, H.; Shafiee, A.

Eur. J. Biochem. 256, 528-534, 1998

A;Title: The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506.

A;Reference number: Z18779; MUID:98451508; PMID:9780228

A;Accession: T17426

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1504 <MOT>

A;Cross-references: EMBL:AF082100; NID:g3798623; PID:g3798625; PIDN:AAC68816.1

A;Experimental source: strain MA6548

C;Genetics:

A;Gene: fkbP

C;Function:

A;Description: required during the biosynthesis of the immunosuppressant FK506 for the activation and incorporation of the pipecolate moiety into the completed acyl chain

C;Superfamily: Mycobacterium tuberculosis mbtE protein; acetate-CoA ligase homology; acyl carrier protein homology

C;Keywords: carrier protein; phosphopantetheine; phosphoprotein

F;533-982/Domain: acetate-CoA ligase homology <ACL>

F;999-1067/Domain: acyl carrier protein homology <ACP>

F;1031/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 51.9%; Score 42; DB 2; Length 1504;  
Best Local Similarity 53.3%; Pred. No. 1.6e+02;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;



Qy 1 DLEMPVLPVEPFPPFV 15  
:| :| | ||| |  
Db 1423 ELRLPGLRTEPFPPV 1437

RESULT 12

B97432

glucose-6-phosphate 1-dehydrogenase (g6pd) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)

C;Species: Agrobacterium tumefaciens

C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002

C;Accession: B97432

R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, B.S.; Cao, Y.; Askenazi, M.; Halling, C.; Mullin, L.; Houmiel, K.; Gordon, J.; Vaudin, M.; Iartchouk, O.; Epp, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Flanagan, C.; Crowell, C.; Gurson, J.; Lomo, C.; Sear, C.; Strub, G.; Cielo, C.; Slater, S.  
Science 294, 2323-2328, 2001

A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens C58.

A;Reference number: A97359; MUID:21608551; PMID:11743194

A;Accession: B97432

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-503 <KUR>

A;Cross-references: GB:AE007869; PIDN:AAK86411.1; PID:g15155545; GSPDB:GN00169

C;Genetics:

A;Gene: AGR\_C\_1065

A;Map position: circular chromosome

C;Superfamily: glucose-6-phosphate dehydrogenase

Query Match 50.6%; Score 41; DB 2; Length 503;  
Best Local Similarity 46.7%; Pred. No. 68;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPFPPFV 15  
|: :||||| |  
Db 12 DMSSQIIPVEPFDCV 26

RESULT 13

T41201

isoleucyl-trna synthetase - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000

C;Accession: T41201

R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Jimenez Martinez, J.  
submitted to the EMBL Data Library, July 1999

A;Reference number: Z21978

A;Accession: T41201

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-973 <WOO>

A;Cross-references: EMBL:AL109736; PIDN:CAB52155.1; GSPDB:GN00068

A;Experimental source: strain 972h-; cosmid c18B5

C;Genetics:

A;Gene: SPBC18B5.08c

A;Map position: 3  
C;Superfamily: isoleucine-tRNA ligase

Query Match 50.6%; Score 41; DB 2; Length 973;  
Best Local Similarity 53.8%; Pred. No. 1.4e+02;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 3 EMPVLPVEPFPPV 15  
| | : | | : | | :  
Db 325 ENPLLPKQSFPPFL 337

RESULT 14

T00333

hypothetical protein KIAA0560 - human

C;Species: Homo sapiens (man)

C;Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 11-Jan-2002

C;Accession: T00333

R;Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.

DNA Res. 5, 31-39, 1998

A;Title: Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.

A;Reference number: Z14086; MUID:98290545; PMID:9628581

A;Accession: T00333

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-1421 <NAG>

A;Cross-references: EMBL:AB011132; NID:d1185402; PIDN:BAA25486.1

A;Experimental source: brain; clone HH1648

C;Genetics:

A;Note: KIAA0560

Query Match 50.6%; Score 41; DB 2; Length 1421;  
Best Local Similarity 50.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 LEMPVLPVEPFP 13  
| : : : | | | |  
Db 1264 LHLHIIPTEPFP 1275

RESULT 15

AG3000

1-acyl-sn-glycerol-3-phosphate acyltransferase plsC [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C;Species: Agrobacterium tumefaciens

C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002

C;Accession: AG3000

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; Kitajima, J.P.; Okura, V.K.; Almeida Jr., N.F.; Zhou, Y.; Bovee Sr., D.; Chapman, P.; Clendenning, J.; Deatherage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClelland, E.; Palmieri, A.; Raymond, C.; Rouse, G.; Saenphimmachak, C.; Wu, Z.; Gordon, D.; Eisen, J.A.; Paulsen, I.; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, B.; Liao, L.; Kim, S.; Hendrick, C.; Zhao, Z.; Dolan, M.; Tingey, S.V.; Tomb, J.; Gordon, M.P.; Olson, M.V.; Nester, E.W.  
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A;Reference number: AB2577; MUID:21608550; PMID:11743193

A;Accession: AG3000

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-266 <KUR>

A;Cross-references: GB:AE008689; PIDN:AAL44421.1; PID:g17742021; GSPDB:GN00187

A;Experimental source: strain C58 (Dupont)

C;Genetics:

A;Gene: plsC

A;Map position: linear chromosome

Query Match 49.4%; Score 40; DB 2; Length 266;

Best Local Similarity 53.8%; Pred. No. 48;

Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEFPF 13

||::||:|| |

Db 174 DLQVPVIPVAMHP 186

Search completed: August 24, 2004, 15:52:51

Job time : 18.5522 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:51:19 ; Search time 54.291 Seconds  
(without alignments)  
86.825 Million cell updates/sec

Title: US-09-641-801-5  
Perfect score: 81  
Sequence: 1 DLEMPVLPVEPFPPFV 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1295152 seqs, 314255058 residues

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	
No.	Score Match Length DB ID	Description

1	81	100.0	15	14	US-10-281-652-5	Sequence 5, Appli
2	49	60.5	104	12	US-10-425-114-58864	Sequence 58864, A
3	48	59.3	232	15	US-10-104-047-3111	Sequence 3111, Ap
4	45	55.6	446	12	US-10-282-122A-77182	Sequence 77182, A
5	44	54.3	135	16	US-10-437-963-108040	Sequence 108040,
6	43	53.1	325	12	US-10-424-599-254766	Sequence 254766,
7	42	51.9	68	12	US-10-424-599-284024	Sequence 284024,
8	42	51.9	174	12	US-10-424-599-204286	Sequence 204286,
9	42	51.9	249	12	US-10-424-599-144697	Sequence 144697,
10	42	51.9	457	15	US-10-369-493-18781	Sequence 18781, A
11	41.5	51.2	86	12	US-10-115-123-456	Sequence 456, App
12	41.5	51.2	86	14	US-10-012-542-456	Sequence 456, App
13	41	50.6	88	12	US-10-424-599-204279	Sequence 204279,
14	41	50.6	101	12	US-10-276-774-1842	Sequence 1842, Ap
15	41	50.6	113	12	US-10-424-599-223803	Sequence 223803,
16	41	50.6	173	12	US-10-424-599-204281	Sequence 204281,
17	41	50.6	175	12	US-10-424-599-204284	Sequence 204284,
18	41	50.6	175	12	US-10-424-599-204285	Sequence 204285,
19	41	50.6	189	16	US-10-437-963-144895	Sequence 144895,
20	41	50.6	308	16	US-10-437-963-171233	Sequence 171233,
21	41	50.6	397	14	US-10-156-761-8560	Sequence 8560, Ap
22	41	50.6	632	16	US-10-437-963-145459	Sequence 145459,
23	41	50.6	644	16	US-10-437-963-112405	Sequence 112405,
24	41	50.6	8026	14	US-10-132-134-12	Sequence 12, Appl
25	40	49.4	97	12	US-10-424-599-216454	Sequence 216454,
26	40	49.4	151	12	US-10-425-114-38710	Sequence 38710, A
27	40	49.4	151	16	US-10-437-963-166092	Sequence 166092,
28	40	49.4	167	16	US-10-437-963-192065	Sequence 192065,
29	40	49.4	219	9	US-09-922-480-2	Sequence 2, Appli
30	40	49.4	219	9	US-09-923-236-2	Sequence 2, Appli
31	40	49.4	219	9	US-09-922-469-2	Sequence 2, Appli
32	40	49.4	219	10	US-09-746-783-106	Sequence 106, App
33	40	49.4	221	14	US-10-231-417-192	Sequence 192, App
34	40	49.4	239	12	US-10-424-599-216458	Sequence 216458,
35	40	49.4	264	16	US-10-437-963-104274	Sequence 104274,
36	40	49.4	318	16	US-10-437-963-137718	Sequence 137718,
37	40	49.4	331	12	US-10-425-114-42069	Sequence 42069, A
38	40	49.4	460	12	US-10-425-114-52303	Sequence 52303, A
39	40	49.4	561	12	US-10-424-599-239201	Sequence 239201,
40	40	49.4	850	15	US-10-369-493-10911	Sequence 10911, A
41	40	49.4	1533	16	US-10-437-963-104266	Sequence 104266,
42	40	49.4	2195	16	US-10-437-963-104271	Sequence 104271,
43	39.5	48.8	233	12	US-10-221-625-65	Sequence 65, Appl
44	39.5	48.8	318	16	US-10-437-963-188608	Sequence 188608,
45	39.5	48.8	361	12	US-10-425-114-41293	Sequence 41293, A

#### ALIGNMENTS

RESULT 1  
 US-10-281-652-5  
 ; Sequence 5, Application US/10281652  
 ; Publication No. US20030091606A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: STANTON, G. John

; APPLICANT: HUGHES, Thomas K.  
 ; APPLICANT: BOLDOGH, Istvan  
 ; TITLE OF INVENTION: USE OF COLOSTRININ, CONSTITUENT PEPTIDES THEREOF, AND  
 ; TITLE OF INVENTION: ANALOGS THEREOF AS OXIDATIVE STRESS REGULATORS  
 ; FILE REFERENCE: 265.00220101  
 ; CURRENT APPLICATION NUMBER: US/10/281,652  
 ; CURRENT FILING DATE: 2002-10-28  
 ; PRIOR APPLICATION NUMBER: US/09/641,803  
 ; PRIOR FILING DATE: 2000-08-17  
 ; PRIOR APPLICATION NUMBER: 60/149,310  
 ; PRIOR FILING DATE: 1999-08-17  
 ; NUMBER OF SEQ ID NOS: 34  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 5  
 ; LENGTH: 15  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
 ; OTHER INFORMATION: peptide  
 US-10-281-652-5

Query Match 100.0%; Score 81; DB 14; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPFPPFV 15  
 |||||  
 Db 1 DLEMPVLPVEPFPPFV 15

# RESULT 2

US-10-425-114-58864  
 ; Sequence 58864, Application US/10425114  
 ; Publication No. US20040034888A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jingdong  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Screen, Steven E  
 ; APPLICANT: Tabaska, Jack E  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated  
 With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53313)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,114  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 58864  
 ; LENGTH: 104  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: 700431848\_FLI.pep  
 US-10-425-114-58864

Query Match 60.5%; Score 49; DB 12; Length 104;  
Best Local Similarity 60.0%; Pred. No. 8.5;  
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPFFV 15  
|| :||| :|| |:  
Db 61 DLWLPLPFPFLFL 75

RESULT 3

US-10-104-047-3111  
; Sequence 3111, Application US/10104047  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3111  
; LENGTH: 232  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-104-047-3111

Query Match 59.3%; Score 48; DB 15; Length 232;  
Best Local Similarity 66.7%; Pred. No. 27;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 EMPVLPVEPFF 14  
: |||| |:  
Db 209 KFPVLPVHPWF 220

RESULT 4

US-10-282-122A-77182  
; Sequence 77182, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A

```

; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77182
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Vibrio cholerae
US-10-282-122A-77182

```

```

Query Match          55.6%; Score 45; DB 12; Length 446;
Best Local Similarity 53.3%; Pred. No. 1.4e+02;
Matches      8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy      1 DLEMPVLPVEPPFPV 15
        |||:| | :| ||::
Db      76 DLEVPNLELEAFPYL 90

```

# RESULT 5

```

US-10-437-963-108040
; Sequence 108040, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B

```



; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 108040  
; LENGTH: 135  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_12332C.1.pep  
US-10-437-963-108040

Query Match 54.3%; Score 44; DB 16; Length 135;  
Best Local Similarity 47.4%; Pred. No. 60;  
Matches 9; Conservative 2; Mismatches 4; Indels 4; Gaps 1;

Qy 1 DLEMPVLPVEP----FPPFV 15  
| ||:|:| ||||  
Db 75 DAPMPEIPIHPPPPVFPPFV 93

RESULT 6

US-10-424-599-254766  
; Sequence 254766, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated  
With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 254766  
; LENGTH: 325  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_72076C.1.pep  
US-10-424-599-254766

Query Match 53.1%; Score 43; DB 12; Length 325;  
Best Local Similarity 77.8%; Pred. No. 2e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 5 PVLPEPFP 13  
||:|| |||  
Db 191 PVVPVHPFP 199

RESULT 7

US-10-424-599-284024  
; Sequence 284024, Application US/10424599  
; Publication No. US20040031072A1

```
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 284024
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_9849C.1.pep
US-10-424-599-284024
```

```
Query Match          51.9%; Score 42; DB 12; Length 68;
Best Local Similarity 53.3%; Pred. No. 59;
Matches      8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
```

```
Qy      1 DLEMPVLPVEPFPPFV 15
          | | | | | :
Db      44 DAEWPCLYKEPFPLI 58
```

#### RESULT 8

US-10-424-599-204286

```
; Sequence 204286, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 204286
; LENGTH: 174
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_26498C.1.pep
US-10-424-599-204286
```

```
Query Match          51.9%; Score 42; DB 12; Length 174;
Best Local Similarity 54.5%; Pred. No. 1.5e+02;
Matches      6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 DLEMPVLPVEP 11  
:::|:|  
Db 77 NIDIPVIPVEP 87

RESULT 9

US-10-424-599-144697  
; Sequence 144697, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated  
With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 144697  
; LENGTH: 249  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_101680C.1.pep  
US-10-424-599-144697

Query Match 51.9%; Score 42; DB 12; Length 249;  
Best Local Similarity 80.0%; Pred. No. 2.2e+02;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 PVLPVEPFPPF 14  
| | | | | | |  
Db 48 PVLPVEGLPF 57

RESULT 10

US-10-369-493-18781  
; Sequence 18781, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION  
OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 18781  
; LENGTH: 457  
; TYPE: PRT  
; ORGANISM: Anabaena PCC7120  
US-10-369-493-18781

Query Match 51.9%; Score 42; DB 15; Length 457;  
Best Local Similarity 63.6%; Pred. No. 4e+02;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 5 PVLPVEPFPPFV 15  
|::||| ||| |  
Db 364 PIMPVMPFPDV 374

RESULT 11

US-10-115-123-456  
; Sequence 456, Application US/10115123  
; Publication No. US20030065151A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: 94 Human Secreted Proteins  
; FILE REFERENCE: PZ029G30AP1D2  
; CURRENT APPLICATION NUMBER: US/10/115,123  
; CURRENT FILING DATE: 2002-04-04  
; PRIOR APPLICATION NUMBER: PCT/US99/13418  
; PRIOR FILING DATE: 1999-06-15  
; PRIOR APPLICATION NUMBER: 60/089,507  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089,508  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089,509  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089,510  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/090,112  
; PRIOR FILING DATE: 1998-06-22  
; PRIOR APPLICATION NUMBER: 60/090,113  
; PRIOR FILING DATE: 1998-06-22  
; NUMBER OF SEQ ID NOS: 532  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 456  
; LENGTH: 86  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-115-123-456

Query Match 51.2%; Score 41.5; DB 12; Length 86;  
Best Local Similarity 47.1%; Pred. No. 89;  
Matches 8; Conservative 3; Mismatches 1; Indels 5; Gaps 1;

Qy 2 LEMPVLP-----VEFPF 13  
||:|:| : |||  
Db 17 LEVPILPTHLLIHPFP 33

RESULT 12

US-10-012-542-456  
 ; Sequence 456, Application US/10012542  
 ; Publication No. US20030044851A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ruben et al.  
 ; TITLE OF INVENTION: 94 Human Secreted Proteins  
 ; FILE REFERENCE: PZ029P1  
 ; CURRENT APPLICATION NUMBER: US/10/012,542  
 ; CURRENT FILING DATE: 2001-12-12  
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325  
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14  
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507  
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16  
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508  
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16  
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509  
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16  
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510  
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16  
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112  
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22  
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113  
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22  
 ; NUMBER OF SEQ ID NOS: 532  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 456  
 ; LENGTH: 86  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-012-542-456

Query Match 51.2%; Score 41.5; DB 14; Length 86;  
 Best Local Similarity 47.1%; Pred. No. 89;  
 Matches 8; Conservative 3; Mismatches 1; Indels 5; Gaps 1;

Qy 2 LEMPVLP-----VEPFP 13  
 ||:||:| : |||  
 Db 17 LEVPILPTHLLIHPFP 33

# RESULT 13

US-10-424-599-204279  
 ; Sequence 204279, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated  
 With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 204279

```
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(88)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_26491C.1.pep
US-10-424-599-204279
```

```
Query Match          50.6%; Score 41; DB 12; Length 88;
Best Local Similarity 45.5%; Pred. No. 1.1e+02;
Matches      5; Conservative      6; Mismatches      0; Indels      0; Gaps      0;
```

```
Qy      1 DLEMPVLPVEP 11
        ::::|::|::|
Db      77 NIDIPVPIEP 87
```

```
RESULT 14
US-10-276-774-1842
; Sequence 1842, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y, Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 1842
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(101)
; OTHER INFORMATION: Xaa = any amino acid or nothing
US-10-276-774-1842
```

```
Query Match          50.6%; Score 41; DB 12; Length 101;
Best Local Similarity 38.5%; Pred. No. 1.2e+02;
Matches      5; Conservative      6; Mismatches      2; Indels      0; Gaps      0;
```

```
Qy      1 DLEMPVLPVEPFP 13
        |:::| :|:: ||
Db      57 DMQVPSVPIQQFP 69
```

```
RESULT 15
```

US-10-424-599-223803  
; Sequence 223803, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated  
With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 223803  
; LENGTH: 113  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_44123C.1.pep  
US-10-424-599-223803

Query Match 50.6%; Score 41; DB 12; Length 113;  
Best Local Similarity 70.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 5 PVLPEPFPPF 14  
|: || ||||  
Db 94 PLGPVSPFPF 103

Search completed: August 24, 2004, 16:41:20  
Job time : 56.291 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:23:00 ; Search time 46.3433 Seconds  
(without alignments)  
102.124 Million cell updates/sec

Title: US-09-641-801-5  
Perfect score: 81  
Sequence: 1 DLEMPVLPVEPFPPV 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriaphage:\*  
17: sp\_archaeophages:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	ID	Description
No.						

---



1	48	59.3	232	4	Q8NAJ2	Q8naj2 homo sapien
2	47	58.0	1766	3	P78584	P78584 aspergillus
3	46	56.8	180	17	Q8TW43	Q8tw43 methanopyru
4	45	55.6	79	11	Q99ME4	Q99me4 rattus norv
5	45	55.6	80	10	Q8LF16	Q8lf16 arabidopsis
6	45	55.6	91	10	Q8S8R5	Q8s8r5 arabidopsis
7	45	55.6	219	5	Q22140	Q22140 caenorhabdi
8	45	55.6	446	16	Q9KTH8	Q9kth8 vibrio chol
9	45	55.6	557	11	Q99L21	Q99l21 mus musculu
10	45	55.6	599	11	Q8C4X8	Q8c4x8 mus musculu
11	45	55.6	599	11	Q8C3K9	Q8c3k9 mus musculu
12	45	55.6	599	11	Q8C2N3	Q8c2n3 mus musculu
13	45	55.6	607	11	Q8BSP2	Q8bsp2 mus musculu
14	45	55.6	619	4	Q9H5Q7	Q9h5q7 homo sapien
15	45	55.6	668	4	Q9H5C7	Q9h5c7 homo sapien
16	45	55.6	772	5	Q9BLH1	Q9blh1 bombyx mori
17	45	55.6	902	16	Q8G762	Q8g762 bifidobacte
18	45	55.6	1307	4	Q9C093	Q9c093 homo sapien
19	45	55.6	1744	11	Q9R095	Q9r095 rattus norv
20	43	53.1	388	5	Q9XTX7	Q9xtx7 caenorhabdi
21	43	53.1	421	17	O28350	O28350 archaeoglob
22	43	53.1	444	10	Q9SR06	Q9sr06 arabidopsis
23	42.5	52.5	681	10	Q9LRV1	Q9lr1 arabidopsis
24	42	51.9	218	3	Q03778	Q03778 saccharomyc
25	42	51.9	232	5	Q9VYG4	Q9vyg4 drosophila
26	42	51.9	260	10	Q94AV8	Q94av8 arabidopsis
27	42	51.9	265	16	Q889N4	Q889n4 pseudomonas
28	42	51.9	367	5	Q22069	Q22069 caenorhabdi
29	42	51.9	448	16	Q8YXW8	Q8yxw8 anabaena sp
30	42	51.9	468	5	Q9NF32	Q9nf32 drosophila
31	42	51.9	469	5	Q9W5D6	Q9w5d6 drosophila
32	42	51.9	473	16	Q8YMB2	Q8ymb2 anabaena sp
33	42	51.9	529	10	Q84WU7	Q84wu7 arabidopsis
34	42	51.9	607	11	Q8CAZ3	Q8caz3 mus musculu
35	42	51.9	696	8	Q85WS8	Q85ws8 pinus korai
36	42	51.9	1312	3	Q8WZV2	Q8wzv2 neurospora
37	42	51.9	1504	2	Q9ZGA6	Q9zga6 streptomyce
38	41	50.6	174	10	Q8RWM7	Q8rwm7 arabidopsis
39	41	50.6	194	16	Q88VZ0	Q88vz0 lactobacill
40	41	50.6	213	10	Q9LVY6	Q9lvy6 arabidopsis
41	41	50.6	309	16	Q89HQ5	Q89hq5 bradyrhizob
42	41	50.6	337	15	Q7SMJ4	Q7smj4 human immun
43	41	50.6	395	16	Q98D83	Q98d83 rhizobium l
44	41	50.6	397	16	Q82PA6	Q82pa6 streptomyce
45	41	50.6	417	2	Q9KWC4	Q9kwc4 agrobacteri

# ALIGNMENTS

## RESULT 1

Q8NAJ2

ID Q8NAJ2 PRELIMINARY; PRT; 232 AA.

AC Q8NAJ2;

DT 01-OCT-2002 (TrEMBLrel. 22, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Hypothetical protein FLJ35269.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Prostate;  
 RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,  
 RA Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,  
 RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,  
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,  
 RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,  
 RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
 RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AK092588; BAC03921.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 232 AA; 25113 MW; 36707CBA84594AC8 CRC64;

Query Match 59.3%; Score 48; DB 4; Length 232;  
 Best Local Similarity 66.7%; Pred. No. 7.6;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 EMPVLPVEPFPPF 14  
 : ||||| |:|  
 Db 209 KFPVLPVHPWPF 220

## RESULT 2

P78584

ID P78584 PRELIMINARY; PRT; 1766 AA.  
 AC P78584;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Polyketide synthase PKS2.  
 GN PKS2.  
 OS Aspergillus parasiticus.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_TaxID=5067;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL 2999;  
 RX MEDLINE=98268975; PubMed=9603849;  
 RA Feng G.H., Leonard T.J.;  
 RT "Culture conditions control expression of the genes for aflatoxin and  
 RT sterigmatocystin biosynthesis in Aspergillus parasiticus and A.  
 RT nidulans.";  
 RL Appl. Environ. Microbiol. 64:2275-2277(1998).  
 DR EMBL; U52151; AAC23536.1; -.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.

DR InterPro; IPR001227; Ac\_trans.  
 DR InterPro; IPR000794; Ketoacyl\_synth.  
 DR InterPro; IPR006163; Pp\_bind.  
 DR Pfam; PF00698; Acyl\_transf; 1.  
 DR Pfam; PF00109; ketoacyl-synt; 1.  
 DR Pfam; PF02801; ketoacyl-synt\_C; 1.  
 DR Pfam; PF00550; pp-binding; 1.  
 DR PROSITE; PS50075; ACP\_DOMAIN; 1.  
 DR PROSITE; PS00606; B\_KETOACYL\_SYNTHASE; 1.  
 KW Phosphopantetheine; Transferase.  
 SQ SEQUENCE 1766 AA; 192068 MW; E20C4BF26F60671E CRC64;

Query Match 58.0%; Score 47; DB 3; Length 1766;  
 Best Local Similarity 60.0%; Pred. No. 89;  
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 DLEMPVLPVEPFPPFV 15  
 |||||: :!  
 Db 1340 DLEMPVLPLATMKYV 1354

# RESULT 3

Q8TW43

ID Q8TW43 PRELIMINARY; PRT; 180 AA.  
 AC Q8TW43;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Uncharacterized membrane protein.  
 GN MK1194.  
 OS Methanopyrus kandleri.  
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;  
 OC Methanopyrus.  
 OX NCBI\_TaxID=2320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
 RX MEDLINE=21927647; PubMed=11930014;  
 RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,  
 RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
 RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,  
 RA Malykh A.G., Koonin E.V., Kozyavkin S.A.;  
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
 RT and monophyly of archaeal methanogens."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).  
 DR EMBL; AE010411; AAM02407.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 180 AA; 19963 MW; 8935B0CADA923F75 CRC64;

Query Match 56.8%; Score 46; DB 17; Length 180;  
 Best Local Similarity 71.4%; Pred. No. 12;  
 Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 LEMPVLPVEPFPPFV 15  
 || |||| | |||  
 Db 29 LECSVLPVPPEPFV 42

# RESULT 4

Q99ME4

ID Q99ME4 PRELIMINARY; PRT; 79 AA.  
AC Q99ME4;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE Thyroid hormone-response protein-1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Xie C., Yang Y., Yang Y., Cai D., Cheng G., Li G., Luo M.;  
RT "Rat thyroid hormone-response gene-1 cloned from brain."  
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF348365; AAK15769.1; -.  
SQ SEQUENCE 79 AA; 8994 MW; F884278152833C09 CRC64;

Query Match 55.6%; Score 45; DB 11; Length 79;  
Best Local Similarity 63.6%; Pred. No. 7.7;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 3 EMPVLPVEFPF 13  
|:||||::| |  
Db 8 EVPVLPPLQPLP 18

# RESULT 5

Q8LF16

ID Q8LF16 PRELIMINARY; PRT; 80 AA.  
AC Q8LF16;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Hypothetical protein.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,  
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;  
RT "Full-length messenger RNA sequences greatly improve genome  
RT annotation."  
RL Genome Biol. 0:0-0(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
RA Feldmann K.;  
RT "Full-Length cDNA from Arabidopsis thaliana."  
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AY085097; AAM61651.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 80 AA; 8671 MW; BB1EF444B1A34E81 CRC64;

Query Match 55.6%; Score 45; DB 10; Length 80;  
Best Local Similarity 77.8%; Pred. No. 7.8;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 5 PVLPVEPFP 13  
||:| ||||  
Db 38 PVIPTPEFP 46

RESULT 6

Q8S8R5

ID Q8S8R5 PRELIMINARY; PRT; 91 AA.  
AC Q8S8R5;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Expressed protein.  
GN AT2G02515.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Lin X., Kaul S., Town C.D., Benito M.-I., Creasy T.H., Haas B.J.,  
RA Wu D., Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,  
RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;  
RT "Arabidopsis thaliana chromosome 2 BAC T8K22 genomic sequence."  
RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AC004136; AAM14912.1; -.  
SQ SEQUENCE 91 AA; 9870 MW; CF45D5B3FB66192D CRC64;

Query Match 55.6%; Score 45; DB 10; Length 91;  
Best Local Similarity 77.8%; Pred. No. 8.9;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 5 PVLPVEPFP 13  
||:| ||||  
Db 49 PVIPTPEFP 57

RESULT 7

Q22140

ID Q22140 PRELIMINARY; PRT; 219 AA.  
AC Q22140;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE T04A8.11 protein.  
GN T04A8.11.  
OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Palmer S.;  
 RL Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology."; Science 282:2012-2018(1998).  
 RL EMBL; Z35663; CAA84730.1; -.  
 DR EMBL; Z35663; CAA84730.1; -.  
 DR PIR; T24429; T24429.  
 DR WormPep; T04A8.11; CE01066.  
 DR GO; GO:0005622; C:intracellular; IEA.  
 DR GO; GO:0005840; C:ribosome; IEA.  
 DR GO; GO:0003735; F:structural constituent of ribosome; IEA.  
 DR GO; GO:0006412; P:protein biosynthesis; IEA.  
 DR InterPro; IPR000114; Ribosomal\_L16.  
 DR Pfam; PF00252; Ribosomal\_L16; 1.  
 DR PRINTS; PR00060; RIBOSOMALL16.  
 SQ SEQUENCE 219 AA; 25360 MW; E22A1E0A573C3FDE CRC64;

Query Match 55.6%; Score 45; DB 5; Length 219;  
 Best Local Similarity 63.6%; Pred. No. 22;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LEMPVLPVEPF 12  
 |::||:| |||  
 Db 21 LKLPVMPAEPF 31

# RESULT 8

## Q9KTH8

ID Q9KTH8 PRELIMINARY; PRT; 446 AA.  
 AC Q9KTH8;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE CapK protein, putative.  
 GN VC0924.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=El Tor N16961 / Serotype O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
RA Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*  
RT *cholerae*.";  
RL Nature 406:477-483(2000).  
DR EMBL; AE004175; AAF94086.1; -.  
DR PIR; H82264; H82264.  
DR TIGR; VC0924; -.  
KW Complete proteome.  
SQ SEQUENCE 446 AA; 50713 MW; 85BDAC396E2EC45D CRC64;

Query Match 55.6%; Score 45; DB 16; Length 446;  
Best Local Similarity 53.3%; Pred. No. 45;  
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPFPPFV 15  
|||:| | :| ||::  
Db 76 DLEVPNLELEAFPYL 90

RESULT 9

Q99L21

ID Q99L21 PRELIMINARY; PRT; 557 AA.  
AC Q99L21;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Similar to hypothetical protein 384D8\_6.  
GN 0610010J20RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; BC003900; AAH03900.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 557 AA; 63153 MW; 742CD81FEA32A7C6 CRC64;

Query Match 55.6%; Score 45; DB 11; Length 557;  
Best Local Similarity 77.8%; Pred. No. 57;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 5 PVLPVEPFPP 13  
||:||||:|  
Db 154 PVVPVEPYL 162

RESULT 10

Q8C4X8

ID Q8C4X8 PRELIMINARY; PRT; 599 AA.  
AC Q8C4X8;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Hypothetical protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK080447; BAC37919.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 599 AA; 68024 MW; 8591FEA89E95ECF3 CRC64;

Query Match 55.6%; Score 45; DB 11; Length 599;  
 Best Local Similarity 77.8%; Pred. No. 61;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 5 PVLPVEPFP 13  
 ||:||||:|  
 Db 196 PVVPVEPYP 204

# RESULT 11

Q8C3K9

ID Q8C3K9 PRELIMINARY; PRT; 599 AA.  
 AC Q8C3K9;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Kidney;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK085584; BAC39479.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 599 AA; 68022 MW; 4332A8427D33B8DA CRC64;

Query Match 55.6%; Score 45; DB 11; Length 599;  
 Best Local Similarity 77.8%; Pred. No. 61;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 5 PVLPVEPFP 13



Db                    ||:||||:|  
196 PVVPVEPYP 204

RESULT 12

Q8C2N3

ID    Q8C2N3            PRELIMINARY;            PRT;    599 AA.  
AC    Q8C2N3;  
DT    01-MAR-2003 (TrEMBLrel. 23, Created)  
DT    01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT    01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE    Hypothetical protein.  
OS    Mus musculus (Mouse).  
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX    NCBI\_TaxID=10090;  
RN    [1]  
RP    SEQUENCE FROM N.A.  
RC    STRAIN=NOD; TISSUE=Thymus;  
RX    MEDLINE=22354683; PubMed=12466851;  
RA    The FANTOM Consortium,  
RA    the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT    "Analysis of the mouse transcriptome based on functional annotation of  
RT    60,770 full-length cDNAs."  
RL    Nature 420:563-573(2002).  
DR    EMBL; AK088294; BAC40265.1; -.  
KW    Hypothetical protein.  
SQ    SEQUENCE    599 AA;    68023 MW;    4332A8427EDA9AB8 CRC64;

Query Match                    55.6%;    Score 45;    DB 11;    Length 599;  
Best Local Similarity    77.8%;    Pred. No. 61;  
Matches        7;    Conservative    2;    Mismatches        0;    Indels        0;    Gaps        0;

Qy                    5 PVLPVEPFP 13  
                     ||:||||:|  
Db                    196 PVVPVEPYP 204

RESULT 13

Q8BSP2

ID    Q8BSP2            PRELIMINARY;            PRT;    607 AA.  
AC    Q8BSP2;  
DT    01-MAR-2003 (TrEMBLrel. 23, Created)  
DT    01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT    01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE    Hypothetical protein.  
OS    Mus musculus (Mouse).  
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX    NCBI\_TaxID=10090;  
RN    [1]  
RP    SEQUENCE FROM N.A.  
RC    STRAIN=C57BL/6J; TISSUE=Forelimb;  
RX    MEDLINE=22354683; PubMed=12466851;  
RA    The FANTOM Consortium,  
RA    the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT    "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
DR EMBL; AK031135; BAC27270.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 607 AA; 68945 MW; 616CA03BDD647852 CRC64;

Query Match 55.6%; Score 45; DB 11; Length 607;  
Best Local Similarity 77.8%; Pred. No. 62;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 5 PVLPVEFPF 13  
||:||||:|  
Db 204 PVVPVEPYP 212

#### RESULT 14

Q9H5Q7

ID Q9H5Q7 PRELIMINARY; PRT; 619 AA.  
AC Q9H5Q7;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Hypothetical protein FLJ23164.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,  
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,  
RA Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,  
RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,  
RA Isogai T., Sugano S.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AK026817; BAB15563.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 619 AA; 70742 MW; 6D5395F9BA1265AF CRC64;

Query Match 55.6%; Score 45; DB 4; Length 619;  
Best Local Similarity 50.0%; Pred. No. 64;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEFPF 14  
|:::| |::||  
Db 386 DIKIPENPLEPLPF 399

#### RESULT 15

Q9H5C7

ID Q9H5C7 PRELIMINARY; PRT; 668 AA.  
AC Q9H5C7;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Hypothetical protein FLJ23577.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,  
 RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,  
 RA Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,  
 RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,  
 RA Isogai T., Sugano S.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AK027230; BAB15700.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 668 AA; 76854 MW; 3F46A3EE71940463 CRC64;

Query Match 55.6%; Score 45; DB 4; Length 668;  
 Best Local Similarity 50.0%; Pred. No. 69;  
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPFPPF 14  
 |:::| |::| |  
 Db 435 DIKIPENPLEPLPF 448

Search completed: August 24, 2004, 15:50:44  
 Job time : 54.3433 secs

OM protein - protein search, using sw model

Run on: August 24, 2004, 14:57:04 ; Search time 8.0597 Seconds  
(without alignments)  
96.908 Million cell updates/sec

Title: US-09-641-801-5  
Perfect score: 81  
Sequence: 1 DLEMPVLPVEPFPFV 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

		%					
Result	Query						
No.	Score	Match	Length	DB	ID	Description	
1	43	53.1	1161	1	KCH2_RABIT	Q8wny2 oryctolagus	
2	42	51.9	696	1	RPOC_PINTH	P52733 pinus thunb	
3	40	49.4	438	1	RGSB_MOUSE	Q9z2h1 mus musculu	
4	40	49.4	491	1	G6PD_RHIME	Q9z3s2 rhizobium m	
5	40	49.4	509	1	DNAA_MYCPA	Q91717 mycobacteri	
6	39	48.1	207	1	YIOR_CVBF	P22654 bovine coro	
7	39	48.1	333	1	YE35_PYRAB	Q9uys2 pyrococcus	
8	39	48.1	438	1	MPK5_HUMAN	Q13163 homo sapien	
9	39	48.1	467	1	IRF6_HUMAN	O14896 homo sapien	
10	39	48.1	467	1	IRF6_MOUSE	P97431 mus musculu	
11	39	48.1	526	1	ACH1_YEAST	P32316 saccharomyc	
12	39	48.1	540	1	GRBE_HUMAN	Q14449 homo sapien	
13	39	48.1	576	1	LE12_RALSO	Q8xs25 ralstonia s	
14	39	48.1	676	1	HUTU_HUMAN	Q96n76 homo sapien	
15	39	48.1	712	1	PPK_SYNPX	Q7u3d7 synechococc	
16	39	48.1	739	1	PPK_MYCLE	O33127 mycobacteri	
17	39	48.1	742	1	PPK_MYCTU	P95111 mycobacteri	

18	39	48.1	1020	1	VP34_CANAL	Q92213	candida alb
19	38.5	47.5	302	1	CASB_MACEU	P28550	macropus eu
20	38.5	47.5	1144	1	FLNC_MOUSE	Q8vbx6	mus musculu
21	38.5	47.5	2725	1	FLNC_HUMAN	Q14315	homo sapien
22	38	46.9	121	1	AMEL_TACAC	O97647	tachyglossu
23	38	46.9	190	1	Y417_ARCFU	O29830	archaeoglob
24	38	46.9	265	1	IHA_SHEEP	P38440	ovis aries
25	38	46.9	477	1	MM03_HUMAN	P08254	homo sapien
26	38	46.9	533	1	LCP2_HUMAN	Q13094	homo sapien
27	38	46.9	582	1	HEM0_OPSTA	P43090	opsanus tau
28	38	46.9	636	1	DXS_SYNLE	Q9r6s7	synechococc
29	38	46.9	636	1	DXS_SYNP7	Q8gaa0	synechococc
30	38	46.9	1135	1	RBL2_RAT	O55081	rattus norv
31	38	46.9	2044	1	SIF2_DROME	P91620	drosophila
32	37.5	46.3	865	1	CPN_DROME	Q02910	drosophila
33	37	45.7	167	1	SERO_GALME	O76192	galleria me
34	37	45.7	239	1	PNUC_SALTY	P24520	salmonella
35	37	45.7	272	1	ATP6_BUCAP	O51878	buchnera ap
36	37	45.7	346	1	XYLD_RHIME	Q92mt4	rhizobium m
37	37	45.7	357	1	REFAK_ECOLI	P27242	escherichia
38	37	45.7	417	1	Y943_METJA	Q58353	methanococc
39	37	45.7	429	1	ISPG_DEIRA	Q9rxc9	deinococcus
40	37	45.7	446	1	PFES_PSEAE	Q04804	pseudomonas
41	37	45.7	486	1	ENV_HTLV2	P03383	human t-cel
42	37	45.7	738	1	PAP_BOVIN	P25500	bos taurus
43	37	45.7	738	1	PAP_MOUSE	Q61183	mus musculu
44	37	45.7	1040	1	RIK1_SCHPO	Q10426	schizosacch
45	37	45.7	1159	1	KCH2_HUMAN	Q12809	homo sapien

# ALIGNMENTS

## RESULT 1

### KCH2\_RABIT

ID KCH2\_RABIT STANDARD; PRT; 1161 AA.  
AC Q8WNY2; O02731; O19119; O97586; Q9TV06;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Potassium voltage-gated channel subfamily H member 2 (Ether-a-go-go  
DE related gene potassium channel 1) (ERG1) (RERG) (ra-erg) (Ether-a-go-  
DE go related protein 1) (Eag related protein 1).  
GN KCNH2 OR ERG.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RA Witchel H.J., Hancox J.C., Levi A.J., Meech R.W.;  
RT "RERG - rabbit ventricular ERG K+ channel subunit."  
RL Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE OF 411-571 FROM N.A. (ISOFORM 1/2).  
RX MEDLINE=97164986; PubMed=9012748;  
RA Wymore R.S., Gintant G.A., Wymore R.T., Dixon J.E., McKinnon D.,

RA Cohen I.S.;  
 RT "Tissue and species distribution of mRNA for the IKr-like K<sup>+</sup> channel,  
 RT erg.";  
 RL Circ. Res. 80:261-268(1997).  
 CC -!- FUNCTION: Pore-forming (alpha) subunit of voltage-gated inwardly  
 CC rectifying potassium channel. Channel properties are modulated by  
 CC cAMP and subunit assembly. Mediates the rapidly activating  
 CC component of the delayed rectifying potassium current in heart  
 CC (IKr) (By similarity).  
 CC -!- SUBUNIT: The potassium channel is probably composed of a homo- or  
 CC heterotetrameric complex of pore-forming alpha subunits that can  
 CC associate with modulating beta subunits. Heteromultimer with  
 CC KCNH6/ERG2, KCNH7/ERG3, KCNE1 and KCNE2 (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q8WNY2-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q8WNY2-2; Sequence=VSP\_000971;  
 CC -!- TISSUE SPECIFICITY: Detected in heart, both in atrium and in left  
 CC ventricle.  
 CC -!- DOMAIN: The segment S4 is probably the voltage-sensor and is  
 CC characterized by a series of positively charged amino acids at  
 CC every third position.  
 CC -!- PTM: Phosphorylated on serine and threonine residues (By  
 CC similarity).  
 CC -!- SIMILARITY: Belongs to the potassium channel family. H (Eag)  
 CC subfamily.  
 CC -!- SIMILARITY: Contains 1 cyclic nucleotide-binding domain.  
 CC -!- SIMILARITY: Contains 1 PAS (PER-ARNT-SIM) dimerization domain.  
 CC -!- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U87513; AAB68612.1; -.  
 DR EMBL; AF068736; AAC99425.1; -.  
 DR EMBL; AF105061; AAD39357.1; -.  
 DR EMBL; U75212; AAC48723.1; -.  
 DR InterPro; IPR000595; cNMP\_binding.  
 DR InterPro; IPR003967; Erg\_channel.  
 DR InterPro; IPR005821; Ion\_trans.  
 DR InterPro; IPR001622; K+channel\_pore.  
 DR InterPro; IPR005820; M+channel\_nlg.  
 DR InterPro; IPR001610; PAC.  
 DR InterPro; IPR000700; PAS-assoc\_C.  
 DR InterPro; IPR000014; PAS\_domain.  
 DR Pfam; PF00027; cNMP\_binding; 1.  
 DR Pfam; PF00520; ion\_trans; 1.  
 DR Pfam; PF00785; PAC; 1.  
 DR PRINTS; PR01470; ERGCHANNEL.

DR SMART; SM00100; cNMP; 1.  
 DR SMART; SM00086; PAC; 1.  
 DR PROSITE; PS00888; CNMP\_BINDING\_1; FALSE\_NEG.  
 DR PROSITE; PS00889; CNMP\_BINDING\_2; FALSE\_NEG.  
 DR PROSITE; PS50042; CNMP\_BINDING\_3; 1.  
 DR PROSITE; PS50112; PAS; 1.  
 DR PROSITE; PS50113; PAC; 1.  
 KW Transport; Ion transport; Ionic channel; Voltage-gated channel;  
 KW Potassium channel; Potassium; Potassium transport; Transmembrane;  
 KW Phosphorylation; Glycoprotein; Multigene family; Alternative splicing.  
 FT DOMAIN 1 405 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 406 426 SEGMENT S1 (POTENTIAL).  
 FT TRANSMEM 453 473 SEGMENT S2 (POTENTIAL).  
 FT DOMAIN 474 497 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 498 518 SEGMENT S3 (POTENTIAL).  
 FT TRANSMEM 523 543 SEGMENT S4 (POTENTIAL).  
 FT DOMAIN 544 549 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 550 570 SEGMENT S5 (POTENTIAL).  
 FT DOMAIN 614 634 SEGMENT H5 (PORE-FORMING) (POTENTIAL).  
 FT TRANSMEM 641 661 SEGMENT S6 (POTENTIAL).  
 FT DOMAIN 662 1161 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 17 88 PAS.  
 FT DOMAIN 92 144 PAC.  
 FT DOMAIN 299 302 POLY-PRO.  
 FT NP\_BIND 744 861 CNMP.  
 FT CARBOHYD 600 600 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 69 85 Missing (in isoform 2).  
 FT /FTid=VSP\_000971.  
 FT CONFLICT 411 411 V -> A (IN REF. 2).  
 FT CONFLICT 445 446 PE -> TD (IN REF. 2).  
 FT CONFLICT 553 553 L -> F (IN REF. 2).  
 FT CONFLICT 561 561 L -> C (IN REF. 2).  
 SQ SEQUENCE 1161 AA; 126961 MW; 79B532B2FFBD9AEB CRC64;

Query Match 53.1%; Score 43; DB 1; Length 1161;  
 Best Local Similarity 77.8%; Pred. No. 40;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 5 PVLPVEPFP 13  
 |:||| |||  
 Db 1095 PLLPVSPFP 1103

## RESULT 2

### RPOC\_PINTH

ID RPOC\_PINTH STANDARD; PRT; 696 AA.  
 AC P52733;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6).  
 GN RPOC1.  
 OS Pinus thunbergii (Green pine) (Japanese black pine).  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.  
 OX NCBI\_TaxID=3350;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95024047; PubMed=7937893;  
 RA Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,  
 RA Sugiura M.;  
 RT "Loss of all ndh genes as determined by sequencing the entire  
 RT chloroplast genome of the black pine *Pinus thunbergii*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).  
 CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription  
 CC of DNA into RNA using the four ribonucleoside triphosphates as  
 CC substrates.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA}(N).  
 CC -!- SUBUNIT: In chloroplasts, the RNA polymerase is composed of four  
 CC subunits: alpha, beta, beta', and beta''.  
 CC -!- SUBCELLULAR LOCATION: Chloroplast.  
 CC -!- SIMILARITY: Belongs to the RNA polymerase beta' chain family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D17510; BAA23472.1; -.  
 DR PIR; T07447; T07447.  
 DR HSSP; Q9KWU6; 1HQM.  
 DR InterPro; IPR000722; RNA\_pol\_A.  
 DR InterPro; IPR007080; RNA\_pol\_Rpb1\_1.  
 DR InterPro; IPR007066; RNA\_pol\_Rpb1\_3.  
 DR InterPro; IPR006592; RNA\_polA\_N.  
 DR Pfam; PF04997; RNA\_pol\_Rpb1\_1; 1.  
 DR Pfam; PF00623; RNA\_pol\_Rpb1\_2; 1.  
 DR Pfam; PF04983; RNA\_pol\_Rpb1\_3; 1.  
 DR SMART; SM00663; RPOLA\_N; 1.  
 KW Transferase; Transcription; DNA-directed RNA polymerase; Chloroplast.  
 SQ SEQUENCE 696 AA; 79805 MW; 722B50492E077A63 CRC64;

Query Match 51.9%; Score 42; DB 1; Length 696;  
 Best Local Similarity 66.7%; Pred. No. 34;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 MPVLPVEPFPPFV 15  
 :|||| || | |  
 Db 285 LPVLPPEPRPIV 296

### RESULT 3

#### RGSB\_MOUSE

ID RGSB\_MOUSE STANDARD; PRT; 438 AA.  
 AC Q9Z2H1;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Regulator of G-protein signaling 11 (RGS11) (Fragment).



GN RGS11.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA He W., Wensel T.G.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Inhibits signal transduction by increasing the GTPase  
 CC activity of G protein alpha subunits thereby driving them into  
 CC their inactive GDP-bound form (By similarity).  
 CC -!- SUBUNIT: HETERODIMER WITH GBETA5 (BY SIMILARITY).  
 CC -!- SIMILARITY: Contains 1 RGS domain.  
 CC -!- SIMILARITY: Contains 1 G protein gamma domain.  
 CC -!- SIMILARITY: Contains 1 DEP domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF061934; AAC70012.1; -.  
 DR HSSP; P49795; 1CMZ.  
 DR MGD; MGI:1354739; Rgs11.  
 DR InterPro; IPR000591; DEP.  
 DR InterPro; IPR001770; G-gamma.  
 DR InterPro; IPR000342; Regl\_Gprotein.  
 DR Pfam; PF00610; DEP; 1.  
 DR Pfam; PF00631; G-gamma; 1.  
 DR Pfam; PF00615; RGS; 1.  
 DR PRINTS; PR01301; RGSPROTEIN.  
 DR ProDom; PD001580; Regl\_Gprotein; 1.  
 DR SMART; SM00049; DEP; 1.  
 DR SMART; SM00224; GGL; 1.  
 DR SMART; SM00315; RGS; 1.  
 DR PROSITE; PS50186; DEP; 1.  
 DR PROSITE; PS50058; G\_PROTEIN\_GAMMA; FALSE\_NEG.  
 DR PROSITE; PS50132; RGS; 1.  
 KW Signal transduction inhibitor.  
 FT NON\_TER 1 1  
 FT DOMAIN 6 81 DEP.  
 FT DOMAIN 193 254 G PROTEIN GAMMA-LIKE.  
 FT DOMAIN 275 390 RGS.  
 SQ SEQUENCE 438 AA; 50430 MW; 5E7CF122CA843EA3 CRC64;

Query Match 49.4%; Score 40; DB 1; Length 438;  
 Best Local Similarity 50.0%; Pred. No. 43;  
 Matches 9; Conservative 3; Mismatches 2; Indels 4; Gaps 1;

Qy 2 LEMPVLPVE----PFPFV 15  
 || |:|:| |||:  
 Db 390 LEEAVIPLETKRWPFPL 407

RESULT 4

G6PD\_RHIME

ID G6PD\_RHIME STANDARD; PRT; 491 AA.  
AC Q9Z3S2;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD).  
GN ZWF OR R00704 OR SMC03070.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99328961; PubMed=10400573;  
RA Willis L.B., Walker G.C.;  
RT "A novel Sinorhizobium meliloti operon encodes an alpha-glucosidase  
RT and a periplasmic-binding-protein-dependent transport system for  
RT alpha-glucosides.";  
RL J. Bacteriol. 181:4176-4184(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=21396507; PubMed=11481430;  
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,  
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,  
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;  
RT "Analysis of the chromosome sequence of the legume symbiont  
RT Sinorhizobium meliloti strain 1021.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
CC -!- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADP(+) = D-glucono-  
CC 1,5-lactone 6-phosphate + NADPH.  
CC -!- PATHWAY: Pentose phosphate pathway; first step.  
CC -!- SIMILARITY: Belongs to the glucose-6-phosphate dehydrogenase  
CC family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF045609; AAD12043.1; -.  
DR EMBL; AL591784; CAC45276.1; -.  
DR HSSP; P11411; 1DPG.  
DR InterPro; IPR001282; G6PD.  
DR Pfam; PF00479; G6PD; 1.  
DR Pfam; PF02781; G6PD\_C; 1.  
DR PRINTS; PR00079; G6PDHDRGNASE.  
DR ProDom; PD001129; G6PD; 1.  
DR TIGRFAMs; TIGR00871; zwf; 1.

DR PROSITE; PS00069; G6P\_DEHYDROGENASE; 1.  
 KW Oxidoreductase; NADP; Glucose metabolism; Complete proteome.  
 FT ACT\_SITE 184 184 BY SIMILARITY.  
 FT CONFLICT 401 401 R -> T (IN REF. 1).  
 SQ SEQUENCE 491 AA; 55301 MW; 0D8B1AFD094E1775 CRC64;

Query Match 49.4%; Score 40; DB 1; Length 491;  
 Best Local Similarity 60.0%; Pred. No. 49;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 6 VLPVEFPFV 15  
 ::||||| :|  
 Db 5 IIPVEPFDYV 14

# RESULT 5

## DNAA\_MYCPA

ID DNAA\_MYCPA STANDARD; PRT; 509 AA.

AC Q9L7L7;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Chromosomal replication initiator protein dnaA.

GN DNAA.

OS Mycobacterium paratuberculosis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI\_TaxID=1770;

RN [1]

RP SEQUENCE FROM N.A.

RA Zhang Q., Kapur V.;

RT "Genomic organization of the Mycobacterium avium subsp.

RT paratuberculosis origin of replication region.";

RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.

CC -!- FUNCTION: Plays an important role in the initiation and regulation  
 CC of chromosomal replication. Binds to the origin of replication; it  
 CC binds specifically double-stranded DNA at a 9 bp consensus (dnaA  
 CC box): 5'-TTATC(C/A)A(C/A)A-3'. DnaA binds to ATP and to acidic  
 CC phospholipids (By similarity).

CC -!- SIMILARITY: Belongs to the dnaA family.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AF222789; AAF33692.1; -.

DR HAMAP; MF\_00377; -; 1.

DR InterPro; IPR003593; AAA\_ATPase.

DR InterPro; IPR001957; Bac\_DnaA.

DR Pfam; PF00308; bac\_dnaA; 1.

DR PRINTS; PR00051; DNAA.

DR SMART; SM00382; AAA; 1.

DR TIGRFAMs; TIGR00362; DnaA; 1.

DR PROSITE; PS01008; DNAA; 1.  
KW DNA replication; DNA-binding; ATP-binding.  
FT NP\_BIND 210 217 ATP (POTENTIAL).  
SQ SEQUENCE 509 AA; 56619 MW; 2472F3F644D34EC9 CRC64;

Query Match 49.4%; Score 40; DB 1; Length 509;  
Best Local Similarity 54.5%; Pred. No. 51;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 3 EMPVLPVEPF 13  
: |: | |||  
Db 109 DAIPPAEPFP 119

#### RESULT 6

##### YIOR\_CVBF

ID YIOR\_CVBF STANDARD; PRT; 207 AA.  
AC P22654;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hypothetical protein in nucleocapsid ORF (IORF).  
OS Bovine coronavirus (strain F15) (BCoV) (BCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
OC Coronaviridae; Coronavirus.  
OX NCBI\_TaxID=11129;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89087718; PubMed=3207501;  
RA Cruciere C., Laporte J.;  
RT "Sequence and analysis of bovine enteritic coronavirus (F15) genome.  
RT I. Sequence of the gene coding for the nucleocapsid protein; analysis  
RT of the predicted protein.";  
RL Ann. Inst. Pasteur Virol. 139:123-138(1988).

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M36656; AAA42759.1; -.  
DR PIR; S06869; S06869.  
DR InterPro; IPR004876; Corona\_nucI.  
DR Pfam; PF03187; Corona\_I; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 207 AA; 23001 MW; A4E5DE61171BAB50 CRC64;

Query Match 48.1%; Score 39; DB 1; Length 207;  
Best Local Similarity 54.5%; Pred. No. 28;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 5 PVLPVEPFPFV 15  
|:| :|| | |  
Db 197 PILAIEPLPLV 207

RESULT 7

YE35\_PYRAB

ID YE35\_PYRAB STANDARD; PRT; 333 AA.  
 AC Q9UYS2;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical protein PYRAB14350 precursor.  
 GN PYRAB14350 OR PAB0953.  
 OS Pyrococcus abyssi.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 OX NCBI\_TaxID=29292;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GE5 / Orsay;  
 RX MEDLINE=22511545; PubMed=12622808;  
 RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,  
 RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,  
 RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;  
 RT "An integrated analysis of the genome of the hyperthermophilic  
 RT archaeon Pyrococcus abyssi.";  
 RL Mol. Microbiol. 47:1495-1512(2003).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AJ248287; CAB50340.1; -.  
 DR PIR; G75055; G75055.  
 DR InterPro; IPR007253; CW\_binding\_2.  
 DR InterPro; IPR000835; HTH\_MarR.  
 DR Pfam; PF04122; CW\_binding\_2; 1.  
 DR PRINTS; PR00598; HTHMARR.  
 KW Hypothetical protein; Transmembrane; Signal; Complete proteome.  
 FT SIGNAL 1 23 POTENTIAL.  
 FT CHAIN 24 333 HYPOTHETICAL PROTEIN PYRAB14350.  
 FT TRANSMEM 232 252 POTENTIAL.  
 SQ SEQUENCE 333 AA; 37598 MW; 5C348C36EBBD6F14 CRC64;

Query Match 48.1%; Score 39; DB 1; Length 333;  
 Best Local Similarity 60.0%; Pred. No. 47;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LEMPVLPVEP 11  
 |::|:| |  
 Db 51 LDIPILPVNP 60

RESULT 8

MPK5\_HUMAN

ID MPK5\_HUMAN STANDARD; PRT; 438 AA.  
 AC Q13163;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Dual specificity mitogen-activated protein kinase kinase 5  
 DE (EC 2.7.1.-) (MAP kinase kinase 5) (MAPKK 5) (MAPK/ERK kinase 5).  
 GN MAP2K5 OR PRKMK5 OR MEK5 OR MKK5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=95279403; PubMed=7759517;  
 RA Zhou G., Bao Z.Q., Dixon J.E.;  
 RT "Components of a new human protein kinase signal transduction  
 RT pathway.";  
 RL J. Biol. Chem. 270:12665-12669(1995).  
 CC -!- FUNCTION: INTERACTS SPECIFICALLY WITH ERK5, AND NOT WITH ANOTHER  
 CC MAP KINASE LIKE P38. IS NOT PHOSPHORYLATED BY RAFA, RAFA OR  
 CC RAFC. MAY INTERACT WITH GTPASES SUCH AS CDC42.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN MANY ADULT TISSUE. ABUNDANT IN  
 CC HEART AND SKELETAL MUSCLE.  
 CC -!- PTM: ACTIVATED BY PHOSPHORYLATION ON SER/THR BY MAP KINASE KINASE  
 CC KINASES (BY SIMILARITY).  
 CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. MAP  
 CC kinase kinase subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U25265; AAA96146.1; -.  
 DR Genew; HGNC:6845; MAP2K5.  
 DR MIM; 602520; -.  
 DR GO; GO:0004672; F:protein kinase activity; TAS.  
 DR GO; GO:0007165; P:signal transduction; TAS.  
 DR InterPro; IPR000270; OPR\_PB1.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR008271; Ser\_thr\_pkin\_AS.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00564; PB1; 1.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR SMART; SM00666; PB1; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase;  
 KW ATP-binding; Phosphorylation.

FT	DOMAIN	166	409	PROTEIN KINASE.
FT	NP_BIND	172	180	ATP (BY SIMILARITY).
FT	BINDING	195	195	ATP.
FT	ACT_SITE	283	283	BY SIMILARITY.
FT	MOD_RES	311	311	PHOSPHORYLATION.
FT	MOD_RES	315	315	PHOSPHORYLATION.
FT	MUTAGEN	195	195	K->M: INACTIVATION.
FT	MUTAGEN	311	311	S->A: INACTIVATION.
FT	MUTAGEN	315	315	T->A: INACTIVATION.
SQ	SEQUENCE	438 AA; 48968 MW; 21246312F1640EE2 CRC64;		

Query Match 48.1%; Score 39; DB 1; Length 438;  
 Best Local Similarity 58.8%; Pred. No. 63;  
 Matches 10; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

Qy 1 DLEMPVLPVEPF--PFV 15  
 | : ||||| | |||  
 Db 367 DEDSPVLPVGEFSEPFV 383

# RESULT 9

## IRF6\_HUMAN

ID IRF6\_HUMAN STANDARD; PRT; 467 AA.  
 AC O14896;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Interferon regulatory factor 6 (IRF-6).  
 GN IRF6.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Grossman A., Mittrucker H.W., Antonio L., Ozato K., Mak T.W.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Grafham D.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnierch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [4]  
 RP VARIANTS VWS VAL-2; ALA-18; MET-18; ALA-39; GLY-61; ARG-70; SER-76;  
 RP HIS-88; GLY-90; HIS-98; GLN-250; ARG-273; 290-PHE--ASP-296 DELINS LEU;  
 RP PRO-294; ILE-297; GLU-320; MET-321; GLU-325; PRO-345; PHE-347;  
 RP SER-369; TRP-374 AND GLU-388, VARIANTS PPS GLY-60; THR-66; LYS-82;  
 RP CYS-84; HIS-84; GLU-89 AND ASN-430, AND VARIANT ILE-274.  
 RX MEDLINE=22242581; PubMed=12219090;  
 RA Kondo S., Schutte B.C., Richardson R.J., Bjork B.C., Knight A.S.,  
 RA Watanabe Y., Howard E., de Lima R.L.L., Daack-Hirsch S., Sander A.,  
 RA McDonald-McGinn D.M., Zackai E.H., Lammer E.J., Aylsworth A.S.,  
 RA Ardinger H.H., Lidral A.C., Pober B.R., Moreno L., Arcos-Burgos M.,  
 RA Valencia C., Houdayer C., Bahuau M., Moretti-Ferreira D.,  
 RA Richieri-Costa A., Dixon M.J., Murray J.C.;  
 RT "Mutations in IRF6 cause Van der Woude and popliteal pterygium  
 RT syndromes.";  
 RL Nat. Genet. 32:285-289(2002).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- DISEASE: Defects in IRF6 are a cause of van der Woude syndrome  
 CC (VWS) [MIM:119300]; also known as lip-pit syndrome (LPS). It is an  
 CC autosomal dominant developmental disorder characterized by lower  
 CC lip pits, cleft lip and/or cleft palate. Penetrance is incomplete.  
 CC Van der Woude and popliteal pterygium syndrome are allelic  
 CC disorders.  
 CC -!- DISEASE: Defects in IRF6 are the cause of popliteal pterygium  
 CC syndrome (PPS) [MIM:119500]. PPS is an autosomal dominant  
 CC developmental disorder characterized by cleft lip and/or cleft  
 CC palate, and skin and genital anomalies. Penetrance is incomplete  
 CC and expressivity is variable. It shows orofacial phenotypic  
 CC similarities with van der Woude syndrome. Van der Woude and  
 CC popliteal pterygium syndrome are allelic disorders.  
 CC -!- SIMILARITY: Belongs to the IRF family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF027292; AAB84111.1; -.  
 DR EMBL; AL022398; CAA18545.1; -.  
 DR EMBL; BC014852; AAH14852.1; -.  
 DR HSSP; P23906; 2IRF.  
 DR TRANSFAC; T05118; -.  
 DR Genew; HGNC:6121; IRF6.  
 DR MIM; 607199; -.  
 DR MIM; 119300; -.  
 DR MIM; 119500; -.



DR InterPro; IPR001346; IRF.  
 DR InterPro; IPR008984; SMAD\_FHA.  
 DR Pfam; PF00605; IRF; 1.  
 DR PRINTS; PR00267; INTERFREGFCT.  
 DR ProDom; PD002355; IRF; 1.  
 DR SMART; SM00348; IRF; 1.  
 DR PROSITE; PS00601; IRF; 1.  
 KW Transcription regulation; DNA-binding; Nuclear protein; Polymorphism;  
 KW Disease mutation.

FT	DNA_BIND	9	111	TRYPTOPHAN PENTAD REPEAT.
FT	VARIANT	2	2	A -> V (in VWS).
FT				/FTId=VAR_014961.
FT	VARIANT	18	18	V -> A (in VWS).
FT				/FTId=VAR_014962.
FT	VARIANT	18	18	V -> M (in VWS).
FT				/FTId=VAR_014963.
FT	VARIANT	39	39	P -> A (in VWS).
FT				/FTId=VAR_014964.
FT	VARIANT	60	60	W -> G (in PPS).
FT				/FTId=VAR_014965.
FT	VARIANT	61	61	A -> G (in VWS).
FT				/FTId=VAR_014966.
FT	VARIANT	66	66	K -> T (in PPS).
FT				/FTId=VAR_014967.
FT	VARIANT	70	70	G -> R (in VWS).
FT				/FTId=VAR_014968.
FT	VARIANT	76	76	P -> S (in VWS).
FT				/FTId=VAR_014969.
FT	VARIANT	82	82	Q -> K (in PPS).
FT				/FTId=VAR_014970.
FT	VARIANT	84	84	R -> C (in PPS).
FT				/FTId=VAR_014971.
FT	VARIANT	84	84	R -> H (in PPS).
FT				/FTId=VAR_014972.
FT	VARIANT	88	88	N -> H (in VWS).
FT				/FTId=VAR_014973.
FT	VARIANT	89	89	K -> E (in PPS).
FT				/FTId=VAR_014974.
FT	VARIANT	90	90	S -> G (in VWS).
FT				/FTId=VAR_014975.
FT	VARIANT	98	98	D -> H (in VWS).
FT				/FTId=VAR_014976.
FT	VARIANT	250	250	R -> Q (in VWS).
FT				/FTId=VAR_014977.
FT	VARIANT	273	273	Q -> R (in VWS).
FT				/FTId=VAR_014978.
FT	VARIANT	274	274	V -> I (common polymorphism; 3% in
FT				European-descended and 22% in Asian
FT				populations).
FT				/FTId=VAR_014979.
FT	VARIANT	290	296	FTSKLLD -> L (in VWS).
FT				/FTId=VAR_014980.
FT	VARIANT	294	294	L -> P (in VWS).
FT				/FTId=VAR_014981.
FT	VARIANT	297	297	V -> I (in VWS).
FT				/FTId=VAR_014982.
FT	VARIANT	320	320	K -> E (in VWS).

```

FT      /FTId=VAR_014983.
FT  VARIANT      321      321      V -> M (in VWS).
FT      /FTId=VAR_014984.
FT  VARIANT      325      325      G -> E (in VWS).
FT      /FTId=VAR_014985.
FT  VARIANT      345      345      L -> P (in VWS).
FT      /FTId=VAR_014986.
FT  VARIANT      347      347      C -> F (in VWS).
FT      /FTId=VAR_014987.
FT  VARIANT      369      369      F -> S (in VWS).
FT      /FTId=VAR_014988.
FT  VARIANT      374      374      C -> W (in VWS).
FT      /FTId=VAR_014989.
FT  VARIANT      388      388      K -> E (in VWS).
FT      /FTId=VAR_014990.
FT  VARIANT      430      430      D -> N (in PPS).
FT      /FTId=VAR_014991.
SQ  SEQUENCE      467 AA;  53129 MW;  7E28F5E0F5BA4053 CRC64;

```

```

Query Match          48.1%;  Score 39;  DB 1;  Length 467;
Best Local Similarity 41.7%;  Pred. No. 67;
Matches      5;  Conservative      5;  Mismatches      2;  Indels      0;  Gaps      0;

```

```

Qy      1 DLEMPVLPVEPF 12
        ::|:| |::||
Db      199 EMEVPQAPIQPF 210

```

# RESULT 10

## IRF6\_MOUSE

```

ID  IRF6_MOUSE      STANDARD;      PRT;      467 AA.
AC  P97431;
DT  28-FEB-2003 (Rel. 41, Created)
DT  28-FEB-2003 (Rel. 41, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Interferon regulatory factor 6 (IRF-6).
GN  IRF6.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BALB/c; TISSUE=Colon;
RA  Grossman A., Mittrucker H.W., Antonio L., Mak T.W.;
RL  Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN  [2]
RP  TISSUE SPECIFICITY.
RX  MEDLINE=22242581; PubMed=12219090;
RA  Kondo S., Schutte B.C., Richardson R.J., Bjork B.C., Knight A.S.,
RA  Watanabe Y., Howard E., de Lima R.L.L., Daack-Hirsch S., Sander A.,
RA  McDonald-McGinn D.M., Zackai E.H., Lammer E.J., Aylsworth A.S.,
RA  Ardinger H.H., Lidral A.C., Pober B.R., Moreno L., Arcos-Burgos M.,
RA  Valencia C., Houdayer C., Bahuau M., Moretti-Ferreira D.,
RA  Richieri-Costa A., Dixon M.J., Murray J.C.;
RT  "Mutations in IRF6 cause Van der Woude and popliteal pterygium
RT  syndromes.";

```

RL Nat. Genet. 32:285-289(2002).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- TISSUE SPECIFICITY: High levels of expression along the medial  
 CC edge of the fusing palate, tooth buds, hair follicles, genitalia  
 CC and skin.  
 CC -!- SIMILARITY: Belongs to the IRF family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U73029; AAB36714.1; -.  
 DR HSSP; P23906; 2IRF.  
 DR TRANSFAC; T05119; -.  
 DR MGD; MGI:1859211; Irf6.  
 DR InterPro; IPR001346; IRF.  
 DR InterPro; IPR008984; SMAD\_FHA.  
 DR Pfam; PF00605; IRF; 1.  
 DR PRINTS; PR00267; INTERNREGFCT.  
 DR ProDom; PD002355; IRF; 1.  
 DR SMART; SM00348; IRF; 1.  
 DR PROSITE; PS00601; IRF; 1.  
 KW Transcription regulation; DNA-binding; Nuclear protein.  
 FT DNA\_BIND 9 111 TRYPTOPHAN PENTAD REPEAT.  
 SQ SEQUENCE 467 AA; 53106 MW; 68CCAA90680FEDC8 CRC64;

Query Match 48.1%; Score 39; DB 1; Length 467;  
 Best Local Similarity 41.7%; Pred. No. 67;  
 Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPF 12  
 :|:| | :|:|  
 Db 199 EMEVPQAPIQPF 210

#### RESULT 11

##### ACH1\_YEAST

ID ACH1\_YEAST STANDARD; PRT; 526 AA.  
 AC P32316;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Acetyl-CoA hydrolase (EC 3.1.2.1) (Acetyl-CoA deacylase) (Acetyl-CoA  
 DE acylase).  
 GN ACH1 OR YBL015W OR YBL0304 OR YBL03.18.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=90237039; PubMed=1970569;  
 RA Lee F.-J.S., Lin L.-W., Smith J.A.;

RT "A glucose-repressible gene encodes acetyl-CoA hydrolase from  
 RT *Saccharomyces cerevisiae*.";  
 RL J. Biol. Chem. 265:7413-7418(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c;  
 RX MEDLINE=93070614; PubMed=1441754;  
 RA van Dyck L., Purnelle B., Skala J., Goffeau A.;  
 RT "An 11.4 kb DNA segment on the left arm of yeast chromosome II  
 RT carries the carboxypeptidase Y sorting gene PEP1, as well as ACH1,  
 RT FUS3 and a putative ARS.";  
 RL Yeast 8:769-776(1992).  
 CC -!- FUNCTION: PRESUMABLY INVOLVED IN REGULATING THE INTRACELLULAR  
 CC ACETYL-COA POOL FOR FATTY ACID AND CHOLESTEROL SYNTHESIS AND  
 CC FATTY ACID OXIDATION. IT MAY BE INVOLVED IN OVERALL REGULATION  
 CC OF ACETYLATION DURING MELATONIN SYNTHESIS.  
 CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O = CoA + acetate.  
 CC -!- SUBUNIT: Monomer.  
 CC -!- PTM: Glycosylated; contains mannose.  
 CC -!- SIMILARITY: TO N.CRASSA ACU-8, AND SOME, TO C.KLUYVERI CAT1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M31036; AAA34388.1; -.  
 DR EMBL; X68577; CAA48570.1; -.  
 DR EMBL; Z35776; CAA84834.1; -.  
 DR PIR; S28549; S28549.  
 DR GermOnline; 138450; -.  
 DR SGD; S0000111; ACH1.  
 DR GO; GO:0005829; C:cytosol; IDA.  
 DR GO; GO:0005739; C:mitochondrion; IDA.  
 DR GO; GO:0003986; F:acetyl-CoA hydrolase activity; IDA.  
 DR GO; GO:0006083; P:acetate metabolism; IMP.  
 DR InterPro; IPR003702; ActCoA\_hydro.  
 DR Pfam; PF02550; AcetylCoA\_hydro; 1.  
 KW Hydrolase; Glycoprotein.  
 FT CONFLICT 308 308 L -> F (IN REF. 2).  
 FT CONFLICT 320 320 S -> A (IN REF. 2).  
 FT CONFLICT 363 364 FP -> LG (IN REF. 2).  
 SQ SEQUENCE 526 AA; 58768 MW; C0C61904F2196A9D CRC64;

Query Match 48.1%; Score 39; DB 1; Length 526;  
 Best Local Similarity 41.2%; Pred. No. 76;  
 Matches 7; Conservative 6; Mismatches 2; Indels 2; Gaps 1;

QY 1 DLEMPVLPV--EPFPFV 15  
 |::||| | :|:|::  
 Db 191 DIDMPVNPFPKPYPYL 207

RESULT 12

GRBE\_HUMAN

ID GRBE\_HUMAN STANDARD; PRT; 540 AA.

AC Q14449;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Growth factor receptor-bound protein 14 (GRB14 adapter protein).

GN GRB14.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96218175; PubMed=8647858;

RA Daly R.J., Sanderson G.M., Janes P.W., Sutherland R.L.;

RT "Cloning and characterization of GRB14, a novel member of the GRB7

RT gene family.";

RL J. Biol. Chem. 271:12502-12510(1996).

CC -!- FUNCTION: Interacts with the cytoplasmic domain of the

CC autophosphorylated insulin receptor which is then inhibited. The

CC interaction is mediated by the SH2 domain (By similarity).

CC -!- SUBUNIT: Binds to the ankyrin repeat region of TNKS2 via its N-

CC terminus.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and

CC endosomes.

CC -!- TISSUE SPECIFICITY: Expressed at high levels in the liver, kidney,

CC pancreas, testis, ovary, heart and skeletal muscle.

CC -!- PTM: Phosphorylated on serine residues.

CC -!- SIMILARITY: Contains 1 PH domain.

CC -!- SIMILARITY: Contains 1 Ras-associating domain.

CC -!- SIMILARITY: Contains 1 SH2 domain.

CC -!- SIMILARITY: Belongs to the GRB7/10/14 family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; L76687; AAC15861.1; -.

DR HSSP; P35235; 1AYA.

DR Genew; HGNC:4565; GRB14.

DR MIM; 601524; -.

DR GO; GO:0005070; F:SH3/SH2 adaptor protein activity; TAS.

DR GO; GO:0007165; P:signal transduction; TAS.

DR InterPro; IPR001849; PH.

DR InterPro; IPR000159; RA\_domain.

DR InterPro; IPR000980; SH2.

DR Pfam; PF00169; PH; 1.

DR Pfam; PF00788; RA; 1.

DR Pfam; PF00017; SH2; 1.

DR PRINTS; PR00401; SH2DOMAIN.

DR ProDom; PD000093; SH2; 1.

DR SMART; SM00233; PH; 1.

DR SMART; SM00314; RA; 1.  
 DR SMART; SM00252; SH2; 1.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 DR PROSITE; PS50200; RA; 1.  
 DR PROSITE; PS50001; SH2; 1.  
 KW SH2 domain; Phosphorylation.  
 FT DOMAIN 106 192 RAS-ASSOCIATING.  
 FT DOMAIN 234 342 PH.  
 FT DOMAIN 439 535 SH2.  
 SQ SEQUENCE 540 AA; 60954 MW; A8FCFC16D7437B47 CRC64;

Query Match 48.1%; Score 39; DB 1; Length 540;  
 Best Local Similarity 50.0%; Pred. No. 78;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 DLEMPVLPVEPFPPF 14  
 ||::| :| | ||  
 Db 69 DLDVPEMPSIPNPF 82

# RESULT 13

## LE12\_RALSO

ID LE12\_RALSO STANDARD; PRT; 576 AA.  
 AC Q8XSZ5;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE 2-isopropylmalate synthase 2 (EC 2.3.3.13) (Alpha-isopropylmalate  
 DE synthase 2) (Alpha-IPM synthetase 2).  
 GN LEUA2 OR RSP0322 OR RS05445.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OG Plasmid megaplasmid.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GMI1000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
 RA Chandler M., Choisine N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 RA Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";  
 RL Nature 415:497-502(2002).  
 CC -!- FUNCTION: Catalyzes the condensation of the acetyl group of  
 CC acetyl-CoA with 3-methyl-2-oxobutanoate (2-oxoisovalerate) to form  
 CC 3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).  
 CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + 3-methyl-2-oxobutanoate + H(2)O =  
 CC 2-hydroxy-2-isopropylsuccinate + CoA.  
 CC -!- PATHWAY: Leucine biosynthesis; first step.  
 CC -!- SUBUNIT: Homotetramer (By similarity).  
 CC -!- SIMILARITY: Belongs to the alpha-IPM synthetase / homocitrate  
 CC synthase family. LeuA 2 subfamily.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AL646078; CAD17473.1; -.  
DR HAMAP; MF\_00572; -, 1.  
DR InterPro; IPR002034; AIPM/Hcit\_synth.  
DR InterPro; IPR000891; HMGL-like.  
DR InterPro; IPR005668; LeuA\_yeast.  
DR Pfam; PF00682; HMGL-like; 1.  
DR TIGRFAMs; TIGR00970; leuA\_yeast; 1.  
DR PROSITE; PS00815; AIPM\_HOMOCIT\_SYNTH\_1; 1.  
DR PROSITE; PS00816; AIPM\_HOMOCIT\_SYNTH\_2; 1.  
KW Leucine biosynthesis; Transferase; Plasmid; Complete proteome.  
SQ SEQUENCE 576 AA; 63149 MW; BBCB0A9A66BA332B CRC64;

Query Match 48.1%; Score 39; DB 1; Length 576;  
Best Local Similarity 66.7%; Pred. No. 84;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 EMPVLPVEP 11  
||| ||::|  
Db 356 EMPYLPIDP 364

#### RESULT 14

##### HUTU\_HUMAN

ID HUTU\_HUMAN STANDARD; PRT; 676 AA.  
AC Q96N76;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Probable urocanate hydratase (EC 4.2.1.49) (Urocanase)  
DE (Imidazolonepropionate hydrolase).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,  
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,  
RA Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,  
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,  
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
RA Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,  
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,  
RA Masuho Y., Nagai K., Isogai T.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.  
CC -!- CATALYTIC ACTIVITY: 3-(5-oxo-4,5-dihydro-3-H-imidazol-4-  
CC yl)propanoate = urocanate + H(2)O.

CC -!- COFACTOR: NAD (By similarity).  
 CC -!- PATHWAY: Histidine degradation; second step.  
 CC -!- SIMILARITY: Belongs to the urocanase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AK055862; BAB71032.1; -.  
 DR GK; Q96N76; -.  
 DR MIM; 276880; -.  
 DR InterPro; IPR000193; Urocanase.  
 DR Pfam; PF01175; Urocanase; 1.  
 DR ProDom; PD025423; Urocanase; 1.  
 DR PROSITE; PS01233; UROCANASE; 1.  
 KW Hypothetical protein; Histidine metabolism; Lyase; NAD.  
 SQ SEQUENCE 676 AA; 74830 MW; C940D3D068648D17 CRC64;

Query Match 48.1%; Score 39; DB 1; Length 676;  
 Best Local Similarity 46.2%; Pred. No. 1e+02;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 DLEMPVLPVEPFP 13  
 |:|| |:| :|  
 Db 82 DIEMRAYPIEQYP 94

# RESULT 15

## PPK\_SYNPX

ID PPK\_SYNPX STANDARD; PRT; 712 AA.  
 AC Q7U3D7;  
 DT 15-MAR-2004 (Rel. 43, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Polyphosphate kinase (EC 2.7.4.1) (Polyphosphoric acid kinase) (ATP-  
 DE polyphosphate phosphotransferase).  
 GN PPK OR SYNW2495.  
 OS Synechococcus sp. (strain WH8102).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
 OX NCBI\_TaxID=84588;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22825697; PubMed=12917641;  
 RA Palenik B., Brahamsha B., Larimer F.W., Land M., Hauser L., Chain P.,  
 RA Lamerdin J., Regala W., Allen E.E., McCarren J., Paulsen I.,  
 RA Dufresne A., Partensky F., Webb E.A., Waterbury J.;  
 RT "The genome of a motile marine Synechococcus."  
 RL Nature 424:1037-1042(2003).  
 CC -!- FUNCTION: Catalyzes the reversible transfer of the terminal  
 CC phosphate of ATP to form a long-chain polyphosphate (polyP).  
 CC -!- CATALYTIC ACTIVITY: ATP + {phosphate}(N) = ADP + {phosphate}(N+1).  
 CC -!- PTM: An intermediate of this reaction is the autophosphorylated  
 CC ppk in which a phosphate is covalently linked to histidine



```

CC      residues through a N-P bond (By similarity).
CC      -!- SIMILARITY: Belongs to the polyphosphate kinase family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; BX569695; CAE09010.1; -.
DR      HAMAP; MF_00347; -; 1.
DR      InterPro; IPR003414; PP_kinase.
DR      Pfam; PF02503; PP_kinase; 1.
KW      Transferase; Phosphorylation; Complete proteome.
FT      ACT_SITE      458      458      PHOSPHOHISTIDINE INTERMEDIATE (BY
FT                                     SIMILARITY).
FT      ACT_SITE      477      477      PHOSPHOHISTIDINE INTERMEDIATE (BY
FT                                     SIMILARITY).
SQ      SEQUENCE      712 AA;  80943 MW;  451977FE71AD95B2 CRC64;

```

```

Query Match          48.1%;  Score 39;  DB 1;  Length 712;
Best Local Similarity 52.9%;  Pred. No. 1.1e+02;
Matches      9;  Conservative      0;  Mismatches      2;  Indels      6;  Gaps      1;

```

```

Qy          5 PVL-----PVEPFPPFV 15
             |||         |  ||||
Db          140 PVLTPLAVDPAHPFPFV 156

```

```

Search completed: August 24, 2004, 15:43:30
Job time : 10.0597 secs

```